AD/A-003 354

JOINT GENERALIZED LEAST SQUARES APPLIED TO COST ESTIMATION FOR FIGHTER AIRCRAFT

Patrick W. O'Brien

Air Force Institute of Technology
Wright-Patterson Air Force Base, Ohio

December 1974

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER GSA/SM/74D-7	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) JOINT GENERALIZED LEAST SQUARES APPLIED TO COST ESTIMATION FOR FIGHTER AIRCRAFT		5. TYPE OF REPORT & PERIOD COVERED AFIT Thesis
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Patrick W. O'Brien Captain USAF		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Air Force Institute of Technology (AU) AFIT/ENS Wright-Patterson Air Force Base OH 45433		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Dean of Engineering (AFIT/EN) Air Force Institute of Technology Wright-Patterson Air Force Base OH 45433		12. REPORT DATE December 1974
		13. NUMBER OF PAGES 102
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE N/A

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Approved for public release; IAW AFR 190-17

JERRY C. HIX, Captain, USAF
Director of Information

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Parametric Cost Estimation
Least Squares
Joint Generalized Least Squares
Fighter Aircraft Cost Estimation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Joint Generalized Least Squares is an extension of least squares techniques which decreases statistical uncertainty in derived regression equations. The technique is applied to historical costs for airframes, avionics, and engines in fighter aircraft.

A comparison is made of parametric cost estimating relationships derived using ordinary and Joint Generalized Least Squares to demonstrate reductions in statistical uncertainty.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

JOINT GENERALIZED LEAST SQUARES

APPLIED TO COST ESTIMATION FOR

FIGHTER AIRCRAFT

THESIS

GSA/SM/74D-7 Patrick W. O'Brien
Captain USAF

JOINT GENERALIZED LEAST SQUARES APPLIED TO COST ESTIMATION FOR FIGHTER AIRCRAFT

THESIS

Presented to the Faculty of the School of Engineering

of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science

by

Patrick W. O'Brien

Captain USAF

Graduate Systems Analysis

December 1974

Approved for public release; distribution unlimited

ib

Preface

Joint Generalized Least Squares is an existing statistical technique that can be used in the development of aircraft cost estimating relationships. The technique represents extensions of Ordinary Least Squares in its theory and assumptions. The Introduction, Section I, and Summary, Section VIII, of this thesis report provide the major emphasis of the research, a summary of the techniques employed, and outline the effect of Joint Generalized Least Squares applied to aircraft cost estimation. These two sections are non-mathematically oriented. A more thorough development of the technique is presented in the remaining sections.

As in any cost analysis, data gathering represented a major portion of the effort. The assistance of personnel from the ASD Comptroller Office, ASD/ACCI, was instrumental in obtaining cost data. Specifically, the assistance of Mr. Paul Shoemaker, Lt. Craig Lentzsch, and Miss Frances Williams is acknowledged and appreciated.

Contents

Contents

List of Tables

Abstract

Joint Genralized Least Squares is a statistical technique which allows the interaction of contemporaneous regression equations, through either related coefficients or correlated disturbances. Aircraft cost estimation is generally accomplished by disaggregation to estimates of three sublevels of total cost for a system: airframes, avionics, and engines. Since aircraft systems are contracted as a total package with obvious interaction among airframes, avionics, and engines towards overall performance, similar interactions in costs can be used to reduce the variability of cost estimates for a given aircraft. Past cost estimating relationships have been developed independently, using Ordinary Least Squares for each of the sublevels of cost, and results have shown high statistical variability. Data for fighter and trainer aircraft contracts are used to demonstrate the effect of using Joint Generalized Least Squares to fit the data and develop cost estimating relationships. A comparison is made to relationships developed using Ordinary Least Squares. Although variance will remain high, substantial reductions over Ordinary Least Squares estimates can be realized through application of Joint Generalized Least Squares.

JOINT GENERALIZED LEAST SQUARES

APPLIED TO COST ESTIMATION FOR

FIGHTER AIRCRAFT

I. Introduction

Aircraft Cost Estimation

Cost estimation for aircraft systems is accomplished before awarding defense contracts, during the development process, and on a continuing basis by Systems Program Offices (SPO's) to maintain control of their respective systems. Agencies actively involved in making cost estimates or those interested in the results of estimates include the Congress, Department of Defense, Systems Command, Logistics Command, Aeronautical Systems Division, SPO's, and defense contractors. Because the specific interest of any or all of these agencies may be in one of several possible elements of overall cost, many categories and sublevels of total program cost are used: direct and indirect costs, engineering costs, manufacturing labor costs, tooling costs, flight test costs, development costs, as well as total program costs. The major subsystems for all military aircraft are airframes, avionics, and engines; Defense Department and service cost analyses have, therefore, concentrated on developing estimating relationships for these three elements of total

program cost. They represent the major portions of overall cost, they are generally separately contracted either directly by the services or subcontracted by a primary contractor, and the operational commands ultimately to use the system require some combination of performance from these three elements of the total system. Ordinary Least Squares is the most common technique used to fit functional variables to the cost observations in developing parametric cost estimating relationships (Ref 8; Ref 12; Ref 14).

Linear Models/Ordinary Least Squares

The technique of Ordinary Least Squares is dependent on the hypothesis of a linear (or log-linear) functional form relating explanatory variables to historical cost observations. The linear model is assumed to be stochastic in that random disturbances are included in the model. The disturbances are usually assumed to have a normal distribution.

For a linear model, Ordinary Least Squares uses differential calculus to minimize the sum of squared residuals, the difference between the observed costs and the estimated costs, in estimating parameters for the model. Several assumptions on the linear model underly the use of Ordinary Least Squares. In the context of aircraft cost estimation, Cost Estimating Relationships

(CER's) developed using Ordinary Least Squares assume independence of the elements of total cost.

Joint Generalized Least Squares

When estimating three elements of total cost--airframes, avionics, and engines--the use of Ordinary Least Squares implies that there is no interaction between the individual equations; CER's are developed separately for each of the elements.

Historical results have shown high variances for each of the individual elements and for total program cost (Ref 2; Ref 14; Ref 16). If the estimates for each of the elements are made jointly and interaction exists among the individual equations, more information can be utilized in estimating the coefficients, thus reducing their statistical variance and improving their predictive ability.

Regressions on costs for individual elements can be related through inter-dependent explanatory variables or the joint probability distribution of the random disturbance terms. Joint Generalized Least Squares (JGLS) utilizes the latter, the interaction of disturbance terms between separate CER's, to relate individual equations.

There are three types of probabilistic interactions that can exist between disturbance terms in the joint estimation of coefficients for airframe, avionics, and engine costs: covariance between observations within each specific type of equation; covariance between types

of equations and disimilar observations, i.e., airframe estimate for the F-100 and avionics estimate for the F-101; and contemporaneous covariance--covariance between types of equations for corresponding observations. The last case, contemporaneous covariance, is assumed to exist when applying Joint Generalized Least Squares, while the first two mentioned are assumed to be zero.

Methodology

To investigate the effect of using Joint Generalized Least Squares in the development of parametric aircraft cost estimates, cost observations for fighter/trainer aircraft, obtained from the Cost Library, Aeronautical Systems Division, will be used as a data base. Statistical cost estimating relationships will be developed using Ordinary Least Squares; using the correlation among the residuals, the same coefficients will be re-calculated employing Joint Generalized Least Squares, and results compared.

Organization of Thesis Report

The source of data and aggregation for analysis is presented in Section II; Section III summarizes the mathematics of Ordinary and Joint Generalized Least Squares; Section IV establishes the statistical criteria used in developing functional relationships and explanatory variables; Section V discusses the specific analyses

accomplished and the CER's developed by Ordinary Least Squares; Section VI provides the analysis and results necessary for the application of Joint Generalized Least Squares; Section VII presents conclusions of the Ordinary Least Squares analyses, a comparison of Ordinary Least Squares and Joint Generalized Least Squares CER's, and discusses improvements and results of joint estimation; finally, Section VIII provides a summary and indicates directions for further investigation in using joint estimation. A separate Appendix is provided to present the data used in each analysis.

II. Data/Data Reduction

Scope

In choosing the data base to be used, it must be emphasized that primary interest is in demonstrating the technique of Joint Generalized Least Squares and its effect on predictive capability rather than development of better cost estimating relationships for any specific class of aircraft or costs. However, actual aircraft cost data and corresponding characteristics will be used to obtain the best CER's possible consistent with standard statistical criteria.

Historically, cost studies have concentrated on only one of the three elements of interest--airframes, avionics, or engines--without regard to their interaction. For that reason, most of the studies do not aggregate the observations by type of aircraft, or type of mission. To limit the problem and to provide a relatively homogeneous data base, only fighter/trainer aircraft will be used.

All data used is unclassified; however, the data is *For Official Use Only*, and will be included as a separate Appendix. The majority of the avionics observations were obtained from aggregation of specific "black box" costs listed in a *Confidential* report and those costs will, therefore, not be included. The aircraft systems

for which data was obtained are the F-100, F-101, F-102, F-104, F-105, F-106, T-38, and F-4.

Source of Data

The primary source of data was the Cost Library, Aeronautical Systems Division (ASD), Wright-Patterson AFB. Specifically, Project Backfill (Ref 3), conducted in 1965 and 1966, contains raw data on total program costs, airframe cost, engine cost, and subsystem cost, including both Government Furnished Equipment and subcontracted equipment.

Data in Project Backfill was compiled from Cost Information Reports (CIR's) provided to the Air Force or the Navy by the major contractors. Costs are provided for each contract within the procurement for a given aircraft system and are listed in contracted-year dollars. Principal cost breakdowns for each contract include direct and indirect costs for the total contract, engineering labor, manufacturing labor, tooling, development support, training, and flight test; also included are total airframe costs, engine costs, and major subsystem costs. Contractor profit is included on all cost observations.

Prior to Fiscal Year 1956, avionics and/or all subcontracted costs were not included as part of the CIR. Since several of the aircraft systems used included contracts prior to 1956, another source was used for the avionics cost observations. RMC Final Report UR-097,

Volume II, *Aircraft Installed Airborne Equipment Configuration and Cost Data*, by Garrett Weinberg and Mary Lee Rech (Ref 17:Vol II), classified *Confidential*, lists avionics costs by AN number (designation of avionics components by function) for specific aircraft in contracted-year dollars. This report uses Project Backfill and RAND Memorandum RM-4851, *An Estimating Relationship for Fighter/Interceptor Avionics System Procurement Costs*, by C. Teng, as sources for aircraft avionics cost. Costs are listed by aircraft model, by fiscal year. Because of the deficiencies in Backfill data prior to FY56 and because the report was consistent with, but more comprehensive than Project Backfill, avionics cost observations from the RMC Report were used.

In addition to costs, Project Backfill includes performance and physical characteristics for each model of aircraft procured. Engine characteristics were obtained from Cost Library files, *Jane's All the World's Aircraft*, and RAND Report R-1017-AFPA/PR, *Measuring Technological Change: Aircraft Turbine Engines*, by Arthur J. Alexander and J. R. Nelson.

Data Reduction/Assumptions

Four analyses were conducted on the data. For convenience and ease in later reference, they will be numbered. The principal characteristics of each are listed below.

Analysis #1. Backfill data only, observations corresponding to each contract, regressions made on adjusted reported subsystem costs, engine costs, and (contract) total program costs minus engine and subsystem costs (assumed to represent airframe costs).

Analysis #2. Backfill and RMC Report data, observations corresponding to each contract, regressions made on Backfill reported airframe and engine costs, and RMC Report avionics costs.

Analysis #3. Backfill data only, observations corresponding to each contract, regressions made on airframe and engine costs.

Analysis #4. Backfill and RMC Report data, observations corresponding to separate aircraft models, regressions made on Backfill airframe and engine costs, and RMC Report avionics costs.

Avionics costs for the F-100A, F-100C, F-101A, F-101C, and F-102A (first two contracts) were not explicitly covered in the RMC Report, nor was Backfill data reliable since the contracts were completed before 1956. In these cases, avionics costs were calculated by summing avionics component average unit costs in the contract year when component costs were listed for other aircraft in the same year.

The index used to adjust to constant dollars was that stated for the aircraft industry as a whole in RAND

Report R-568-PR, *Aerospace Price Indexes*, by H. G. Campbell (Ref 4:18). 1970 is used as the base year. For the first analysis, subsystem costs were adjusted to compensate for a lack of consistency in reported costs. Reported subsystem costs, by function, lacked continuity: some specific subsystem costs per unit varied by as much as a factor of four between contracts while others were almost constant. Since the subsystem packages, as a whole, appeared to be relatively homogeneous for a given aircraft system, an average subsystem cost per contract was calculated. Reported subsystem costs per contract were adjusted to constant dollars using the aircraft industry index; the total subsystem cost for each aircraft system was then determined and the average cost per aircraft for each total system procurement calculated; costs per contract in constant dollars were then determined using the quantity per contract; finally, the subsystem costs were "returned" to contract-year dollars using the industry index.

For the first three analyses, Ordinary Least Squares regressions were accomplished on contract-year costs, and the aircraft industry index included as an explanatory variable. For the last analysis, any given aircraft model covered a procurement period of several fiscal years. The costs for each model were computed, then, by adjusting the contract costs to 1970 dollars and adding the costs for each model; adjustment to 1970

dollars was made with the industry wide index for each aspect of cost.

Reliability of Data

Airframe and engine costs reported in Project Backfill include contractor profit, but represent the actual dollars budgeted in the contract year, and, as such, are assumed to be accurate. Subsystem costs reported in Project Backfill listed no cost for Fire Control Systems for some aircraft. Fire Control includes the radar and missile firing control computers, and all but one system, the T-38, carry such systems. The RMC Report includes the manufacturer's reported price for most relevant fiscal years, or at least the development cost of all avionics components; while the validity of the reported costs is open to some question, the costs used are realistic approximations to actual avionics costs.

III. Mathematical Development

This section will provide a brief overview of the mathematical development of Linear Models, Ordinary Least Squares, and Joint Generalized Least Squares. The discussion presented here is based primarily on Theil (Ref 15:Chap 3,6,7), with additional reference to Mendenhall and Schaeffer (Ref 13:Chap 11).

Notation

The most convenient means of discussing linear models and least squares techniques is through the use of vector/matrix notation. The following conventions will be used.

Capital letters designate vectors or matrices. Subscripted lower case letters indicate components of vectors or matrices.

A "hat" over a variable, parameter, etc., indicates an estimate or prediction; for example, the estimate for the vector Y is designated $\hat{Y}$. The notation $\bar{Y}$ indicates a vector the same size as the vector Y, and whose components are all equal to the average of the components of Y.

The transpose of a matrix is indicated by an apostrophe, i.e., X' indicates the transpose of the matrix X. X^{-1} represents the inverse of the matrix X.

Linear Models

A linear model, as the name implies, relates a dependent variable to a linear function of independent variables. Define the following:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} \qquad X = \begin{bmatrix} x_{11} & x_{12} & \cdot & \cdot & \cdot & x_{1k} \\ x_{21} & x_{22} & \cdot & \cdot & \cdot & x_{2k} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ x_{n1} & x_{n2} & \cdot & \cdot & \cdot & x_{nk} \end{bmatrix}$$

$$B = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_k \end{bmatrix} \qquad E = \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_n \end{bmatrix}$$

Y represents a vector of n response observations; X represents the corresponding n observations of k explanatory variables; B represents the vector of k coefficients of the explanatory variables; and E represents a vector of n random disturbance terms.

A linear statistical model relating Y to X is of the form

$$Y = XB + E$$

where the vector of coefficients, B, is unknown, and the expected value of the random disturbance terms is assumed to be zero. Y is not a linear function of the observed matrix, X, but rather a linear function of the unknown vector of coefficients, B. There exist several techniques that can be used to estimate this vector of coefficients, including maximum likelihood and least squares.

Ordinary Least Squares

Least squares is a method of fitting a response defined by a linear function of variables to a set of actual observations of the dependent variable; it represents a technique of estimating the vector of coefficients above.

For a linear model, once the vector of coefficients has been estimated, by whatever method, the vector of estimated responses is given by

$$\hat{Y} = X\hat{B}$$

Ordinary Least Squares minimizes the sum of squared deviations of components of the observed and estimated response vectors. In matrix notation, the value to be minimized is

$$(Y-\hat{Y})'(Y-\hat{Y})$$

Substituting for $\hat{Y}$, this value becomes

$$(Y-X\hat{B})'(Y-X\hat{B})$$

To minimize this function of $\hat{B}$, the derivative with respect

to $\hat{B}$ is calculated, and set equal to zero. First and second order conditions are used to show that the squared residuals are minimized. Solving the first order equations, the Ordinary Least Squares estimated vector of coefficients is given by

$$\hat{B} = (X'X)^{-1}X'Y$$

The principal assumptions made in deriving the statistical properties of Ordinary Least Squares estimates are:

1. The observed elements of the Y vector and the X matrix are measured without error.
2. The n-element random disturbance vector, E, is distributed normally.
3. The conditional mean vector of E given X is

$$\mathcal{E}(E|X) = \emptyset$$

4. The variance-covariance matrix of Y given X is

$$V(Y|X) = \sigma^2 I$$

where σ^2 is an unknown positive parameter and I is the n x n Identity matrix.

Using these assumptions, the following properties of Ordinary Least Squares estimates have been derived (Ref 13:399):

1. The least squares estimates for the coefficient vector are unbiased, i.e., $\mathcal{E}(\hat{B}) = B$.

2. The variance-covariance matrix for the estimated coefficients is given by

$$V(\hat{B}) = \sigma^2 (X'X)^{-1} .$$

3. The components of $\hat{B}$ are each normally distributed.

4. The unbiased estimator for σ^2, commonly denoted S^2, is given by

$$S^2 = \frac{(Y-\hat{Y})'(Y-\hat{Y})}{n-k} .$$

Joint Generalized Least Squares

In many situations, least squares techniques are applied simultaneously to develop several individual equations. These equations may, in fact, be related, but separate equations derived by Ordinary Least Squares implies no interaction, either in the development of individual equations or possible combinations of the equations.

To take advantage of any interactions between equations, joint estimation techniques can be used. In this context, the assumptions underlying Ordinary Least Squares are restrictive. The relation $\sigma^2 I$ of the fourth assumption implies that the combined observations for least squares estimation would have equal variance; in addition, interaction is assumed not to exist since covariance (off-diagonal elements) are zero.

To allow a more accurate representation of the combined observations, the fourth assumption listed for Ordinary Least Squares is restated as

4a. The variance-covariance matrix of Y given X is

$$V(Y|X) = \alpha^2 V$$

where α^2 is an unknown positive parameter and V is a known symmetric positive definite n x n matrix.
This assumption allows for unequal diagonal elements (variances) as well as positive or negative covariance between observations.

To estimate the vector of coefficients for the combined equations, the theory of Ordinary Least Squares is used after transforming the combined observation matrix, (Y X), so that the variance-covariance matrix will be of the form $\alpha^2 I$ rather than $\alpha^2 V$. Since V is symmetric and positive definite, so is its inverse; therefore, there is a non-singular n x n matrix P such that

$$P'P = V^{-1}$$

Premultiplying the combined model, Y = XB + E, by P:

$$PY = PXB + PE$$

the transformed observation matrix is then (PY PX), and the variance-covariance matrix of PY given PX is given by

$$V(PY|PX) = \alpha^2 I$$

The estimated vector of coefficients, after the transformation, is found in the same manner as was used for Ordinary Least Squares. The vector of coefficients, designated B_J, is estimated by

$$\hat{B}_J = (X'V^{-1}X)^{-1}X'V^{-1}Y$$

A more detailed and rigorous development is provided in Theil (Ref 15:236-241). Theil has derived the following properties of the estimated coefficients:

1. The estimates for the coefficient vector are unbiased, i.e., $E(\hat{B}_J) = B_J$.
2. The variance-covariance matrix for the estimated coefficients is given by

$$V(\hat{B}_J) = \alpha^2 (X'V^{-1}X)^{-1}.$$

3. The components of B_J are each normally distributed.
4. An unbiased estimator of α^2 is given by

$$\hat{\alpha}^2 = \frac{(Y-X\hat{B}_J)'V^{-1}(Y-X\hat{B}_J)}{n-k}$$

The estimated vector of coefficients, $\hat{B}_J$, is called the Generalized Least Squares estimator because of the more general assumptions underlying the model. The technique of Joint Generalized Least Squares (JGLS) derives its name through the use of this estimator and the specific form of the matrix, V, that it utilizes.

JGLS uses the correlation and associated covariance between random disturbance terms to relate equations which have been combined for joint estimation. In combining equations, the applicable situation for the use of JGLS is that in which there exists correspondence between observations among the separate equations. For this

situation, there are three types of covariances between disturbance terms in the joint estimation of coefficients: covariance between observations within each specific type of equation; covariance between types of equations and dissimilar observations; and contemporaneous covariance--covariance between types of equations for corresponding observations. The last case, contemporaneous covariance, is assumed to be present, while the first two mentioned are assumed to be zero.

The existence of covariance between similar observations indicates the manner in which the observation matrices will be combined. To demonstrate, assume there are three equations to be combined. The corresponding observation matrices will be designated $(Y_1 \quad X)$, $(Y_2 \quad W)$, and $(Y_3 \quad Z)$ to avoid triple subscripts. Define the following:

$$y_1 = \begin{bmatrix} y_{11} \\ y_{21} \\ \cdot \\ \cdot \\ \cdot \\ y_{n_1 1} \end{bmatrix} \qquad y_2 = \begin{bmatrix} y_{12} \\ y_{22} \\ \cdot \\ \cdot \\ \cdot \\ y_{n_2 2} \end{bmatrix} \qquad y_3 = \begin{bmatrix} y_{13} \\ y_{23} \\ \cdot \\ \cdot \\ \cdot \\ y_{n_3 3} \end{bmatrix}$$

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdot & \cdot & \cdot & x_{1k_1} \\ x_{21} & x_{22} & \cdot & \cdot & \cdot & x_{2k_1} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ x_{n_1 1} & x_{n_1 2} & \cdot & \cdot & \cdot & x_{n_1 k_1} \end{bmatrix}$$

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdot & \cdot & \cdot & w_{1k_2} \\ w_{21} & w_{22} & \cdot & \cdot & \cdot & w_{2k_2} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ w_{n_2 1} & w_{n_2 2} & \cdot & \cdot & \cdot & w_{n_2 k_2} \end{bmatrix}$$

$$Z = \begin{bmatrix} z_{11} & z_{12} & \cdot & \cdot & \cdot & z_{1k_3} \\ z_{21} & z_{22} & \cdot & \cdot & \cdot & z_{2k_3} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ z_{n_3 1} & z_{n_3 2} & \cdot & \cdot & \cdot & z_{n_3 k_3} \end{bmatrix}$$

There are n_1 observations and k_1 explanatory variables for the first equation; n_2 and k_2, n_3 and k_3, for the second and third equations, respectively. Initially assume that

$n_1=n_2=n_3$, and that there is a correspondence between all observations. The combined observation matrix will be of the following form, letting $n=n_1=n_2=n_3$:

$$\begin{bmatrix}
y_{11} & x_{11} & x_{12} \cdots x_{1k_1} & 0 & 0 \cdots 0 & 0 & 0 \cdots 0 \\
y_{12} & 0 & 0 \cdots 0 & w_{11} & w_{12} \cdots w_{1k_2} & 0 & 0 \cdots 0 \\
y_{13} & 0 & 0 \cdots 0 & 0 & 0 \cdots 0 & z_{11} & z_{12} \cdots z_{1k_3} \\
y_{21} & x_{21} & x_{22} \cdots x_{2k_1} & 0 & 0 \cdots 0 & 0 & 0 \cdots 0 \\
y_{22} & 0 & 0 \cdots 0 & w_{21} & w_{22} \cdots w_{2k_2} & 0 & 0 \cdots 0 \\
y_{23} & 0 & 0 \cdots 0 & 0 & 0 \cdots 0 & z_{21} & z_{22} \cdots z_{2k_3} \\
\cdot & \cdot & \cdot \quad \cdot & \cdot & \cdot \quad \cdot & \cdot & \cdot \quad \cdot \\
\cdot & \cdot & \cdot \quad \cdot & \cdot & \cdot \quad \cdot & \cdot & \cdot \quad \cdot \\
\cdot & \cdot & \cdot \quad \cdot & \cdot & \cdot \quad \cdot & \cdot & \cdot \quad \cdot \\
y_{n1} & x_{n1} & x_{n2} \cdots x_{nk_1} & 0 & 0 \cdots 0 & 0 & 0 \cdots 0 \\
y_{n2} & 0 & 0 \cdots 0 & w_{n1} & w_{n2} \cdots w_{nk_2} & 0 & 0 \cdots 0 \\
y_{n3} & 0 & 0 \cdots 0 & 0 & 0 \cdots 0 & z_{n1} & z_{n2} \cdots z_{nk_3}
\end{bmatrix}$$

If Y and X are now defined to indicate this matrix, the vector of estimated coefficients, B_J, is calculated from

$$B_J = (X'V^{-1}X)^{-1}X'V^{-1}Y$$

where B_J is composed of $k_1 + k_2 + k_3$ elements. Assumption 4a states that the V matrix is known, and indicates the variance-covariance of Y given X; unfortunately, the matrix is rarely known. However, provided Ordinary Least Squares estimates have been calculated for the individual

equations, the matrix can be approximated for the application of JGLS. Only contemporaneous covariance is assumed to exist; the variance for each of the equations can be estimated using Ordinary Least Squares techniques. The "allowed" variance-covariance of Y given X in the combined observation matrix above is, therefore, represented by a block-diagonal matrix with n diagonal 3 x 3 submatrices, all equal to the matrix whose diagonal elements are the variance estimates derived from the properties of Ordinary Least Squares for each of the individual equations. The off-diagonal elements of the submatrix are found in the following way. Simple correlation coefficients for the residual vectors, $(Y_i-\hat{Y}_i)$, i = 1, 2, 3, of the separate Ordinary Least Squares equations are calculated. Using the relation between correlation coefficients, r_{ij}, variance estimates, S_i^2, and covariance, S_{ij},

$$r_{ij} = \frac{S_{ij}}{(S_i^2 S_j^2)^{\frac{1}{2}}}$$

the estimates for $S_{12} = S_{21}$, $S_{13} = S_{31}$, and $S_{23} = S_{32}$, are calculated. Define the symmetric 3 x 3 submatrix, D, as

$$D = \begin{bmatrix} S_1^2 & S_{12} & S_{13} \\ S_{21} & S_2^2 & S_{23} \\ S_{31} & S_{32} & S_3^2 \end{bmatrix}$$

The diagonal elements of D are unbiased estimates of the variance of the individual equations, while covariance is represented through the correlation of residuals among the equations.

Approximations for the V matrix and its inverse are

$$\hat{V} = \begin{bmatrix} D & & & & \\ & D & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & D \end{bmatrix} \qquad \hat{V}^{-1} = \begin{bmatrix} D^{-1} & & & & \\ & D^{-1} & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & D^{-1} \end{bmatrix}$$

When the observation matrices for separate equations are not of the same size, the application of JGLS is modified as follows: for corresponding observations, the combined matrix is structured as before; remaining observations are added below these as single rows in the combined matrix. Variances for the separate equations are estimated using all available observations; simple correlation coefficients between residuals are calculated using only corresponding observations; covariances are then calculated as before from the relation between the correlation coefficients, variances, and covariances. The approximation for the V matrix is modified to represent covariance only for corresponding observations. Additional entries along the diagonal of V are scalars representing the variance of "extra" observations:

$$\hat{V} = \left[\begin{array}{ccccccc} D & & & & & & \\ & \cdot & & & & & \\ & & \cdot & & & & \\ & & & \cdot & & & \\ & & & & D & & \\ \hdashline & & & & s^2 & & \\ & & & & & \cdot & \\ & & & & & & \cdot \\ & & & & & & \cdot \\ & & & & & & s^2 \end{array}\right]$$

where D represents the 3 x 3 variance-covariance matrix of corresponding observations, and S^2 represents the variance of individual observations (and may include more than one type). The size of $\hat{V}$ is $(n_1+n_2+n_3)$ by $(n_1+2_n+n_3)$.

In summary, the application of JGLS utilizes the results of Ordinary Least Squares and a specific form of the theory of Generalized Least Squares to provide joint estimates for related equations.

IV. Statistical Criteria

In the development of CER's for each analysis, the following criteria were used to establish functional form, linear or log-linear, and explanatory variables: coefficient of determination, R^2; coefficient of determination adjusted for degrees of freedom, $\bar{R}^2$; correlation between all possible explanatory variables and relevant costs; and t-ratios of coefficients obtained from Ordinary Least Squares fits on the data. All of the statistics used are standard measures of the adequacy of regression developed functional relationships. Detailed explanations of each can be found in most statistical and econometric textbooks (Ref 10; Ref 13; Ref 15). Only a brief description will be presented here, based on Theil (Ref 15; Chap 3, Chap 4), to provide continuity and establish the specific procedures used to analyze the data.

For a linear model, $Y = YB + E$, with n observations and k explanatory variables, for which estimates of the coefficient vector, B, have been made using least squares, let $\hat{B}$ represent the vector of estimated coefficients, $\bar{Y}$ represent an n x 1 vector, all of whose components equal the average of the components of the Y vector, and $\hat{Y} = X\hat{B}$ represent the estimate for the Y vector. Using this model each of the statistics will be discussed separately below.

Coefficient of Determination

The coefficient of determination, R^2, is used as a measure of the degree of the total variability existing in a given set of observations that is explained, or accounted for, by the explanatory variables. As indicated by its notation, R^2 is a positive number, and is defined as the ratio of explained sum of squares to total sum of squares, where both explained and total sum of squares represent squared deviations from their respective means. Notationally, the definition most generally used is in the form of one minus the ratio of unexplained sum of squares to total sum of squares to avoid introducing further matrix notation. This definition is presented below; for the derivation and proof of the equation, see Theil (Ref 15; 175-178).

$$R^2 = 1 - \frac{(Y-\hat{Y})'(Y-\hat{Y})}{(Y-\bar{Y})'(Y-\bar{Y})}$$

From the definitions and assumptions of the Ordinary Least Squares technique and the coefficient of determination, it has been shown that R^2 can have values from zero to one. Since it is obviously desirable to have the unexplained variation as small as possible, "good" values for R^2 are close to one. R^2 will equal one when $Y = \hat{Y}$, or when there is no unexplained variation; R^2, therefore, can be arbitrarily increased by adding explanatory variables (up to the number of observations), but with a subsequent loss of degrees of freedom.

Coefficient of Determination Adjusted for Degrees of Freedom

Although R^2 can be made arbitrarily close to one, the subsequent effect on the unbiased estimate for the variance of the vector of disturbance terms, E, is opposite in nature. The variance estimate, $\hat{\sigma}^2$, is defined as

$$\hat{\sigma}^2 = \frac{(Y-\hat{Y})'(Y-\hat{Y})}{n-k}$$

which is indeterminate when k is equal to n (although $\hat{Y}$ will tend to Y). Obviously, a statistical model would not be developed with k equal to n; however, in the same sense, even when k is not "very small" compared to the number of observations, the term $(Y-\hat{Y})'(Y-\hat{Y})$ in the definition of R^2 will tend to be small and hence give an overly optimistic view of the performance of the explanatory variables.

To take into account the effect of degrees of freedom and better evaluate the performance of explanatory variables, the coefficient of determination adjusted for degrees of freedom is defined as follows:

$$\bar{R}^2 = 1 - \frac{\frac{1}{n-k}(Y-\hat{Y})'(Y-\hat{Y})}{\frac{1}{n-1}(Y-\bar{Y})'(Y-\bar{Y})}$$

Using the definition of R^2 and through algebraic manipulation, the usual form relating $\bar{R}^2$ to R^2 is

$$\bar{R}^2 = R^2 - \frac{k-1}{n-k}(1-R^2)$$

Notice that $\bar{R}^2$ is less than R^2 in all cases except when k equals one or R^2 equals one.

Correlation Analysis

If X_i represents the i^{th} column vector from the X matrix of explanatory variables, and $\bar{X}_i$ is defined as an n x 1 vector whose components are all equal to the average of the components of X_i, the simple correlation coefficients, $r(X_i,Y)$ and $r(X_i,X_j)$, are defined as follows:

$$r(X_i,Y) = \frac{(X_i-\bar{X}_i)'(Y-\bar{Y})}{((X_i-\bar{X}_i)'(X_i-\bar{X}_i)(Y-\bar{Y})'(Y-\bar{Y}))^{\frac{1}{2}}}$$

$$r(X_i,X_j) = \frac{(X_i-\bar{X}_i)'(X_j-\bar{X}_j)}{((X_i-\bar{X}_i)'(X_i-\bar{X}_i)(X_j-\bar{X}_j)'(X_j-\bar{X}_j))^{\frac{1}{2}}}$$

Simple correlation coefficients provide an indication of the degree of the relationship between two vectors of observed responses/variables. Simple correlation coefficients can have values from -1 to +1; high correlation, either negative or positive, is indicated by values near these two extremes.

Simple correlation for any two vectors may be high due to a direct relationship between them, or by the fact that both are linearly related to another set of variables (Ref 15:171). To establish the correlation between two vectors when the effect of all other existing variables in the analysis are removed (holding the remaining variables

constant), partial correlation coefficients are used. Partial correlation coefficients are, in effect, computed by linearly expressing the two relevant variables in terms of the remaining group of variables, using least squares; the simple correlation of the vectors of residuals from these two "fits" then represents their partial correlation coefficient (Ref 15:171-172). Johnston (Ref 10: 132-135) has shown that partial correlation coefficients can be calculated from the co-factors of the matrix of simple correlation coefficients, defined as the k+1 x k+1 matrix whose i,j^{th} element is the simple correlation coefficient between vectors i and j in the n x k+1 matrix (Y X), $i,j = 1,\ldots,k+1$. The matrix of simple correlation coefficients is symmetric since $r(X_i,X_j) = r(X_j,X_i)$, and all diagonal elements are equal to 1. Using the definition of matrix inverse and the form of the partial correlation coefficients developed by Johnston, partial correlation coefficients can be calculated as follows:

let R_s represent the matrix of simple correlation coefficients

$C = R_s^{-1}$ represent the inverse of the matrix R_s

c_{ij} represent the i,j^{th} element of C

r^p_{ij} represent the partial correlation coefficient between vectors i and j, $i \neq j$

Using the notation above, the partial correlation coefficient between two vectors, i and j, is defined to be

$$r^{p}_{ij} = \frac{-c_{ij}}{(c_{ii}c_{jj})^{\frac{1}{2}}}$$

In developing CER's for each analysis, a combination of simple and partial correlation coefficients was used to determine the explanatory variables to be used. In addition, simple correlation coefficients between explanatory variables were used to determine the existence of multicollinearity.

T-ratios of Coefficients

Under the assumptions of Ordinary Least Squares, the random disturbance vector is distributed normally. From the form of the linear model, it then follows that the Y vector is also distributed normally. Finally, since the vector of estimated coefficients, $\hat{B}$, calculated by least squares, is a linear function of the Y vector, $\hat{B}$ is distributed normally (Ref 13:389-399).

If $\hat{b}_i$ represents the i^{th} component of the vector of estimated coefficients, $\hat{B}$, the ratio of $\hat{b}_i$ to the square root of its variance possesses a Student t distribution for a null hypothesis that $b_i = 0$, with n-k degrees of freedom. This "standard" t-ratio is used to evaluate the significance of estimated coefficients in the final CER's.

Methodology

In each of the four analyses, CER's were determined by fitting cost observations to all possible explanatory

variables, using Ordinary Least Squares. From this fit, the highest attainable R^2 was determined. Partial correlation coefficients were evaluated to eliminate variables with low correlation; simple correlation coefficients between explanatory variables were used to eliminate some variables logically similar to others even though partial correlation may have been high to avoid problems of multicollinearity. Variables were eliminated as long as $\bar{R}^2$ increased, while t-ratios for coefficients were evaluated to indicate values significantly different from zero at the 95 percent confidence level.

V. Ordinary Least Squares Cost Estimating Relationships

In analyzing the cost data, several aggregations of cost and assumptions could be made. Four analyses of the data were completed, using historical cost studies and logical assumptions as a basis for establishing functional form, independent cost observations, and eliminating data points.

Explanatory Variables

The aircraft physical and performance characteristics used are listed by aircraft model in Table I. When contracts included more than one model for a particular aircraft, the characteristics for the most advanced model were used, except in cases where exact numbers of each model produced were given. In these cases, if one model was produced in significantly greater quantity than another, its performance characteristics were used.

Engine characteristics are listed in Table II, along with the using aircraft. Several aircraft employed more than one type engine for a given model. In most cases, only one type engine was used per contract; when more than one engine type was procured for a given contract, characteristics for the most advanced engine were used.

One qualitative variable was used to indicate prototype contracts, when these contracts were included in

Table I

Aircraft Physical/Performance Characteristics

Aircraft	AMPR Weight	GTO* Weight	Cmbt** Weight	Max Speed	Cmbt** Speed	Cmbt** Radius	Wing Loading
F-100A	12118	28899	25607	742	579	311	75.2
F-100C(1)	12958	32615	27587	803	785	510	84.8
F-100C(2)	12940	32536	27413	713	655	510	84.5
F-100D	13907	34050	28847	790	632	464	85.2
F-100F	14900	34692	29503	790	632	452	86.8
F-101A	14865	48001	39495	873	863	588	130.4
F-101B	15734	51724	40853	950	530	503	123.5
F-101C	14865	48908	40429	872	856	571	132.9
F/TF-102A	12034	28150	25262	677	553	335	40.4
F-104A	7631	22614	17768	1150	1070	350	115.0
F-104B	8653	22104	17518	1150	975	188	112.5
F-104C	7966	22410	19470	1150	1150	306	114.0
F-104D	8556	21535	18420	1150	1150	157	110.0
F-104F	8761	21735	18640	1150	1150	150	115.0
F-105B	19427	52000	34870	1195	750	566	122.1
F-105D	19082	52838	35637	1192	726	543	127.2
F-105F	21011	54580	37416	1110	714	444	131.8
F-106A(3)	15074	39195	31480	1153	588	426	50.0
T-38A	5246	11761	10233	713	650	320	69.2
F-4B	18400	54600	37500	1220	1220	900	93.0

(1) Aircraft equipped with J-57-21 engines
(2) Aircraft equipped with J-57-39 engines
(3) Characteristics used for the F-106A, although each contract included procurement of F-106B's with slightly different characteristics due to an increased avionics package.

AMPR Weight represents airframe empty weight without installed equipment, including wheel assemblies, engines, and avionics. The weight is defined in the *Aircraft Manufacturer's Planning Report*, hence the name AMPR.

*Gross Take-off

**Combat

(Ref 3)

Table II

Engine Characteristics

Engine	Using Aircraft	TTWR*	Max Thrust (000 lb)	Specific Fuel Consumption	Comp** Ratio
J-57-39	F-100A/C	1.91	14.8	2.1	11.3
J-57-21	F-100C/D/F	1.98	16.0	2.1	11.3
J-57-13	F-101A/C	2.02	16.0	2.1	11.3
J-57-55	F-101B	2.05	16.9	2.3	12.0
J-57-23A	F-102A	2.02	16.0	2.1	11.3
J-79-3A	F-104A/B	2.95	14.8	2.272	12.0
J-79-7	F-104C/D/F	2.96	15.8	1.97	12.2
J-75-5	F-105B	2.5	24.5	2.15	11.9
J-75-19W	F-105D/F	2.71	24.5	2.15	11.9
J-75-17	F-106A/B	2.71	24.5	2.15	11.9
J-85-5A	T-38A	4.77	3.85	2.2	7.0
J-79-2A	F-4A/B	3.13	16.15	2.0	12.0
J-79-8	F-4B	2.72	17.0	1.93	12.9

*Thrust-to-weight ratio. Calculated as the ratio of static military thrust to engine weight.

**Compression

(Ref 2; Ref 3; Ref 9:498-504,509-512)

the analysis. The variable has values of one for prototype contracts, zero otherwise. The other qualitative variable used was number of engines per aircraft.

Finally, quantity per contract, cumulative quantity by aircraft system (as opposed to model of aircraft), and a price index, listed in Table III, were included as explanatory variables.

As stated previously, an aggregate price index for the aircraft industry is used; more specific indices of cost components exist, but since the objective is joint estimation, an overall index seems more relevant. With the exception of the last analysis, when contract costs were combined, the index was not used to adjust the cost data; by including the index as an explanatory variable, a more thorough analysis of the effect of price escalation could be made.

No attempt was made to analyze the effect of learning commonly attributed to the aircraft industry; nor were cumulative quantity observations adjusted using techniques of learning curve analysis to reflect that cumulative quantity which represented the average cost for a given contract. Cumulative quantity, as used, is the sum of the current contract quantity and all previous quantities for that aircraft system.

Table III

Price Adjustment Index

Year	Adjustment Factor	Year	Adjustment Factor
1952	2.882*	1961	1.688
1953	2.706	1962	1.610
1954	2.535	1963	1.545
1955	2.282	1964	1.494
1956	2.158	1965	1.406
1957	2.122	1966	1.311
1958	1.959	1967	1.237
1959	1.838	1968	1.166
1960	1.758	1969	1.076
		1970	1.000

*1952 factor was not given; value was calculated using the average annual rate for the subsequent five years.

(Ref 4:18)

Form of Analyses

For convenience and later reference, the analyses will be numbered in the order accomplished. The form and details of each analysis will be discussed separately.

All regression analysis on the data was accomplished using OMNITAB II, Version 5, dated 15 May 1971, on the Logistics Command General Electric/Honeywell 600 Series computer system.

Analysis #1

The objective in this analysis was to utilize as much of the information in Project Backfill as possible in demonstrating the effects of joint estimation. As discussed in Section II, subsystem costs were adjusted,

using the price index, to compensate for a lack of consistency. Subsystem average unit costs per contract, for similar aircraft models, varied significantly, even when adjusted to constant dollars. Although this variance might be attributed to modifications, retrofits, etc., no characteristics attributable to each subsystem could be found.

Although total reported airframe costs per contract were given, the costs used were calculated by subtracting reported engine costs and the adjusted subsystem costs from total contract cost. This was done to reflect the fact that the airframe (prime) contractor includes costs of mating all subsystems, engines, avionics, landing gear, etc., to the airframe in each contract.

Of the 53 contracts acquired from Project Backfill, 51 observations were used in developing the airframe CER, 45 for subsystems, and 50 for engines. Two contracts were eliminated, both for procurement of the F-100F. These two contracts were the last completed at the Los Angeles plant. The representation of airframe costs as total contract cost minus engine and subsystem cost implies inclusion of the majority of applied contractor overhead to airframes; for these contracts, the overhead per aircraft per month was twice that of any other production contract, possibly because of closing production. In any case, no such tendency for applied overhead to increase was found for

other systems. Six contracts were made previous to 1956 and could not be used in the development of the subsystem CER, due to the lack of reported costs. One contract, for the YF-104A, did not list a separate engine cost. The contract was included as a prototype for airframes, but could not be used for engines. The other contracts included as prototypes are: F-100A (203 aircraft), which, although a production contract, was ordered into production before prototype testing had been made; T-38A (6 aircraft, 4 of which were static, non-flying test aircraft); F-4A (7 aircraft).

In each equation, the dependent variable is average unit cost per contract. Initial regressions on airframe costs included both linear and log-linear functional forms. Linear regressions resulted in an R^2 below .5 and were eliminated from consideration. The log-linear form was also used for the engine and subsystem CER's.

Analysis #2

In this analysis, regressions were made on reported airframe and engine costs from Project Backfill, and avionics costs from RMC Report UR-097, to demonstrate joint estimation, by contract, with the best available information for airframes, avionics, and engines. Because of limited avionics data, and the difference in avionics packages and airframe costs between prototype and production aircraft,

strictly prototype contracts were not used. In addition, the first two F-101A contracts (31 and 84 aircraft) and the first F-4A/B production contract (16 aircraft) were eliminated, because avionics costs could not be determined. Only observations corresponding to avionics costs were used for airframes and engines; a more complete analysis for airframe and engine costs was accomplished in Analysis #3.

After the deletions, 45 observations were used in developing CER's for each of the aspects of cost. Regressions were made for both linear and log-linear functional forms, using total contract costs as dependent variables, as well as regressions on average unit cost per contract. The coefficient of determination, R^2, remained high, above .8, for both airframes and engines, for all regressions; R^2 for avionics was highest using a linear form with contract cost as the dependent variable. Correlation among the residuals for the three costs was extremely low (below .1) for the long-linear form, indicating that little improvement could be expected from joint estimation. For these reasons, and to investigate the effect of linear functional forms, the analysis was completed using contract costs and linear functional forms. Despite the multiplicative nature of the price index, it was included as a linear explanatory variable, rather than adjust the raw data.

One of the advantages of a log-linear model over a linear functional form is that all predicted costs will be positive if the explanatory variables are all positive. The disadvantage of linear models is graphically illustrated in this analysis: six predicted costs are negative, five avionics costs and one airframe cost. A possible solution to this problem is constrained regression analysis, restricting the predictions to positive values; however, since the ultimate objective of the analysis is to show the effect of joint estimation rather than to develop realistic CER's, no adjustment or further analysis was accomplished.

Analysis #3

Because of the nature of the data in Project Backfill it was felt that the most reliable data was for airframes and engines; both costs are explicitly reported for all contracts. The previous analyses required elimination of data points to compensate for the lack of avionics data. Therefore, an analysis was made without avionics or subsystem costs.

Two of the 53 cost observations were eliminated, the YF-100A and YF-104A. Both these contracts were for two test vehicles. The other prototype contracts were included because the nature of the system or the number procured indicated some resemblance to production contracts

for airframe and engine costs; a qualitative variable was included in the analysis, equal to one for the first contract of each aircraft system, zero otherwise, to compensate for higher costs expected in initial contracts. The qualitative variable was also set equal to one for the seventh F-100 observation since it was the first contract completed at a different production facility (Columbus, Ohio) than the previous six (seven including the test vehicle contract).

Regressions were made on average unit cost per contract using a log-linear functional form, which provided the highest R^2 for both aspects of cost.

Analysis #4

Throughout the first three analyses, costs for each contract were used as independent observations, which is consistent with past RAND airframe cost studies (Ref 12). Cost studies for engines and avionics have generally used costs by aircraft model, either total or per unit cost, as independent cost observations; in addition, RAND Working Note, WN-8729-PR, *Cost Estimating Relationships for Aircraft Airframes*, by D. J. Dreyfuss, H. E. Boren, Jr., and G. W. Corwin, indicates a trend for airframe analysis to follow this same approach. Cost analysis along these lines is intuitively appealing since characteristics, and hence, available explanatory variables, are generally

specific to a given model of aircraft; also, analysis by contract partially masks true average costs per unit (by model) because of production quantities for a given contract, variable inflation factors within a procurement cycle, and multiple model contracts.

The objectives of the final analysis were to investigate CER's for independent observations by aircraft model and demonstrate the effect of joint estimation. To establish total costs of each model, contract costs were adjusted to constant dollars using the price index in Table III (page 36); like model aircraft costs were then summed to arrive at total cost for each aspect of cost by aircraft model; using the total number produced, an average unit cost by model was established. Prototype models were not included in the analysis. The following aircraft models comprised independent observations: F-100C/D/F, F-101A/B, F-102A, F-104A/B/C/D/F, F-105B/D/F, F-106A, T-38A, and F-4B. Regressions were made on average unit costs using a log-linear form.

Coefficients of determination were lower in this analysis than in any of the first three; however, residuals and the added sum of squared residuals for all three aspects of cost were comparable with previous analyses using a log-linear form.

Results from Ordinary Least Squares

A summary of the CER's and relevant statistics is presented in Tables IV through VII. Values given in the tables for log-linear forms represent coefficients and statistics for the natural logarithms of costs and the explanatory variables. For these analyses, the CER's are listed in multiplicative form, and represent the estimating relationships for actual costs.

For several of the CER's, the constant (intercept) term was statistically insignificant and was dropped from the relationship. The coefficient of determination, R^2, may have a different interpretation under these circumstances, since the definition of R^2 given in Section III is based on the existence of a constant term; however, to allow comparison between equations, the definition for R^2 in Section III is used in all cases.

At the 95 percent confidence level, there are four coefficients whose t-ratios indicate values not statistically significant. In two cases, the coefficients were for the price index (Tables IV and V, t-ratios of -1.96 and -1.30). Since no adjustment was made to costs for price escalation in Analyses 1, 2 and 3, the price index was retained regardless of significance. In Table VII, the t-ratio for Wing Loading in the airframe CER is -1.57; however, removing the variable resulted in a drop in R^2 from .62 to .48, and the variable was therefore retained.

For the same reason, the coefficient for Wing Loading in the avionics CER, Analysis #2, with a t-ratio of -1.83 (Table V) was included (R^2 decrease from .59 to .37).

In Table VII, the avionics CER contains two speeds as explanatory variables. Although this would seem to indicate multicollinearity, the simple correlation between the variables was reasonably low (less than .5); in addition, with reference to Table I, it can be seen that, in general, interceptor aircraft, the F-101B, F-102, and F-106, have lower combat speeds. Since avionics costs are higher for these aircraft, combat speed appears to compensate for this effect, since the coefficient has a negative sign.

Table IV

Cost Estimating Relationships
Analysis #1

Coefficients

Type of Cost	Explanatory Variable	Estimated Coefficient	t-ratio
Airframes	Cum. Quantity (Q_c)	-.469013	-12.11
	AMPR Weight (W_a)	.783941	5.65
	Max Speed (S)	.416076	2.47
	Prototype Qual. Var. (V)	.580173	3.08
	Price Index (I)	-.608023	-1.96
Avionics	Constant (K)	-20.352315	-9.76
	GTO Wt-AMPR Wt (W_{ga})	.887182	5.80
	Max Speed (S)	3.320458	8.55
	Wing Loading (L)	-1.197235	-5.45
	Price Index (I)	-1.925900	-4.29
Engines	Compression Ratio (R)	3.413712	9.90
	Thrust to Wt Ratio (T_w)	.652594	3.53
	Max Thrust (M)	-.651833	-4.15
	Cum. Quantity (Q_c)	-.162780	-7.33
	Price Index (I)	-.937022	-3.11

Statistics

Equation	R^2	$\bar{R}^2$	$S^2 = \frac{(Y-\hat{Y})'(Y-\hat{Y})}{n-k}$
Airframes	.913	.905	.0956
Avionics	.875	.863	.1744
Engines	.805	.787	.0414

Cost Estimating Relationships

Dependent variables: average unit cost per contract (000$)

Airframes: $C_1 = Q_c^{-.469} W_a^{.784} S^{.416} V^{.580} I^{-.608}$

Avionics: $C_2 = \exp(-20.352) W_{ga}^{.887} S^{3.320} L^{-1.197} I^{-1.926}$

Engines: $C_3 = R^{3.414} T_w^{.653} M^{-.652} Q_c^{-.163} I^{-.937}$

Table V

Cost Estimating Relationships
Analysis #2

Coefficients

Type of Cost	Explanatory Variable	Estimated Coefficient	t-ratio
Airframes	Quantity per contract (Q_d)	.455526	12.62
	Cumulative Quantity (Q_c)	-.037473	-4.96
	Price Index (I)	-14.221773	-2.12
	AMPR Weight (W_a)	.006640	7.28
Avionics	Quantity per contract (Q_d)	.121589	4.93
	Price Index (I)	-24.423455	-4.15
	Max Speed (S)	.056956	3.24
	Wing Loading (L)	-.239384	-1.83
	GTO Wt-AMPR Wt (W_{ga})	.001322	3.04
Engines	Constant (K)	-128.784430	-3.66
	Quantity per contract (Q_d)	.149527	15.58
	Price Index (I)	-6.170923	-1.30
	Specific Fuel Consump. (F)	40.034522	2.81
	Compression Ratio (R)	4.261417	4.54
	Number of Engines (N)	12.096858	3.49

Statistics

Equation	R^2	$\bar{R}^2$	$S^2 = \frac{(Y-\hat{Y})'(Y-\hat{Y})}{n-k}$
Airframes	.833	.821	975.704
Avionics	.589	.548	418.779
Engines	.878	.862	71.988

Cost Estimating Relationships

Dependent variable: cost per contract (millions $)

Airframes: $C_1 = .456Q_d - .037Q_c - 14.222I + .007W_a$

Avionics: $C_2 = .122Q_d - 24.423I + .057S - .239L + .001W_{ga}$

Engines: $C_3 = -128.784 + .15Q_d - 6.171I + 40.035F + 4.261R + 12.097N$

Table VI

Cost Estimating Relationships
Analysis #3

Coefficients			
Type of Cost	Explanatory Variable	Estimated Coefficient	t-ratio
Airframes	Cum. Quantity (Q_c)	-.430131	-15.07
	Price Index (I)	-1.006188	-3.60
	Prototype Qual. Var. (V)	.394893	3.08
	AMPR Weight (W_a)	1.044991	37.63
Engines	Cumulative Quantity (Qc)	-.201026	-9.53
	Price Index (I)	-.772518	-3.20
	Thrust to Wt Ratio (T_w)	.585593	3.76
	Max Thrust (M)	-.685595	-5.23
	Compression Ratio (R)	3.517864	12.18

Statistics

Equation	R^2	$\bar{R}^2$	$S^2 = \frac{(Y-\hat{Y})'(Y-\hat{Y})}{n-k}$
Airframes	.889	.882	.0750
Engines	.864	.852	.0308

Cost Estimating Relationships

Dependent variables: average unit cost per contract (000$)

Airframes: $C_1 = Q_c^{-.430} I^{-1.006} V^{.395} W_a^{1.045}$

Engines: $C_2 = Q_c^{-.201} I^{-.773} T_w^{.586} M^{-.686} R^{3.518}$

Table VII

Cost Estimating Relationships
Analysis #4

Coefficients

Type of Cost	Explanatory Variable	Estimated Coefficient	t-ratio
Airframes	Quantity per model (Q_d)	-.230464	-3.02
	Gross Take-off Weight (W_g)	1.057095	6.17
	Wing Loading (L)	-.525572	-1.57
Avionics	Quantity per model (Q_d)	-.271108	-3.42
	Max Speed (S)	3.063690	5.37
	Combat Speed (S_c)	-2.050794	-3.61
Engines	Constant (K)	14.145672	4.47
	Quantity per model (Q_d)	-.162844	-3.38
	Gross Take-off Weight (W_g)	-1.082210	-2.42
	Max Thrust (M)	1.423641	2.83

Statistics

Equation	R^2	$\bar{R}^2$	$S^2 = \frac{(Y-\hat{Y})'(Y-\hat{Y})}{n-k}$
Airframes	.622	.568	.1402
Avionics	.698	.655	.2396
Engines	.743	.683	.0649

Cost Estimating Relationships

Dependent variables: average unit cost per model (000$)

Airframes: $C_1 = Q_d^{-.230} W_g^{1.057} L^{-.526}$

Avionics: $C_2 = Q_d^{-.271} S^{3.064} S_c^{-2.051}$

Engines: $C_3 = \exp(14.146) Q_d^{-.163} W_g^{-1.082} M^{1.424}$

VI. Application of Joint Generalized Least Squares

The information needed for the application of Joint Generalized Least Squares is discussed in Section III. To summarize, Ordinary Least Squares regressions provide variance estimates and the simple correlation coefficients of residual vectors for corresponding observations are calculated. Using this information, contemporaneous covariance between equations is calculated. An approximation of the variance-covariance matrix, V, for the combined observations, is then constructed. The V matrix is block diagonal, where the submatrices along the diagonal each consist of the variance-covariance matrix for the contemporaneous observations. Finally, the JGLS vector of estimated coefficients, $\hat{B}_J$, is computed from

$$\hat{B}_J = (X'V^{-1}X)^{-1}X'V^{-1}Y$$

A summary of the elements needed for the application of JGLS is given in Tables VIII through XI. Separate tables are provided for each of the four analyses conducted. The V matrix is specified by listing the elements of its diagonal submatrix, D, along with the variance elements, if applicable, for observations with no contemporaneous counterparts (Analysis #1, Table VIII).

In Analysis #1, there were 45 corresponding observations for airframe, avionics (subsystem), and engine costs.

The six additional airframe and five additional engine observations were added to the combined observation matrix as single observations, as discussed in Section III, and assumed to have no covariance, although five of the six additional airframe observations were, in fact, contemporaneous with the extra engine observations. The covariance of these terms was not included because of the nature in which airframe costs were derived, by subtracting reported subsystem and engine costs from total contract cost. Since the six additional airframe observations were those prior to FY56 (no reported subsystem costs), these costs did not have the same relation to engine costs.

The coefficients derived by JGLS and their respective t-ratios are given in Tables XII through XV. For ease in presentation, the coefficients are referenced by providing the Ordinary Least Squares estimates. Further analysis/improvements of JGLS will be discussed in Section VII.

Table VIII

Elements of JGLS
Analysis #1

Variance Estimates From
Ordinary Least Squares

Airframes: $S_1^2 = .0956$

Avionics: $S_2^2 = .1744$

Engines: $S_3^2 = .0414$

Matrix of Simple Correlation Coefficients
For Ordinary Least Squares Residual Vectors

	Airframes	Avionics	Engines
Airframes	1.0000	.1587	.2607
Avionics	.1587	1.0000	.0725
Engines	.2607	.0725	1.0000

Contemporaneous Covariance

Airframes-Avionics: $S_{12} = .0205$

Airframes-Engines: $S_{13} = .0164$

Avionics-Engines: $S_{23} = .0062$

Elements of

$\hat{V}$ $\hat{V}^{-1}$

$$D = \begin{bmatrix} .0956 & .0205 & .0164 \\ .0205 & .1744 & .0062 \\ .0164 & .0062 & .0414 \end{bmatrix} \qquad D^{-1} = \begin{bmatrix} 11.4644 & -1.2925 & -4.3629 \\ -1.1925 & 5.8887 & -.4095 \\ -4.3629 & -.4095 & 25.9442 \end{bmatrix}$$

$S_1^2 = .0956$ $(S_1^2)^{-1} = 10.4603$

$S_3^2 = .0414$ $(S_3^2)^{-1} = 24.1546$

Table IX

Elements of JGLS
Analysis #2

Variance Estimates From Ordinary Least Squares

Airframes: $S_1^2 = 975.7037$

Avionics: $S_2^2 = 418.7794$

Engines: $S_3^2 = 71.9881$

Matrix of Simple Correlation Coefficients For Ordinary Least Squares Residual Vectors

	Airframes	Avionics	Engines
Airframes	1.0000	.4872	.4201
Avionics	.4872	1.0000	.3815
Engines	.4201	.3815	1.0000

Contemporaneous Covariance

Airframes-Avionics: $S_{12} = 311.4289$

Airframes-Engines: $S_{13} = 111.3376$

Avionics-Engines: $S_{23} = 66.2396$

Elements of

$\hat{V}$ $\hat{V}^{-1}$

$$D = \begin{bmatrix} 975.70 & 311.43 & 111.34 \\ 311.43 & 418.78 & 66.24 \\ 111.34 & 66.24 & 71.99 \end{bmatrix} \qquad D^{-1} = \begin{bmatrix} .00147 & -.00086 & -.00148 \\ -.00086 & .00330 & -.00171 \\ -.00148 & -.00171 & .01775 \end{bmatrix}$$

Table X

Elements of JGLS
Analysis #3

Variance Estimates From
Ordinary Least Squares

Airframes: $S_1^2 = .0750$

Engines: $S_2^2 = .0308$

Matrix of Simple Correlation Coefficients
For Ordinary Least Squares Residual Vectors

	Airframes	Engines
Airframes	1.0000	.1377
Engines	.1377	1.0000

Contemporaneous Covariance

Airframes-Engines: $S_{12} = .00661$

Elements of

$\hat{V}$	$\hat{V}^{-1}$
$D = \begin{bmatrix} .0750 & .0066 \\ .0066 & .0308 \end{bmatrix}$	$D^{-1} = \begin{bmatrix} 13.6004 & -2.9259 \\ -2.9259 & 33.1291 \end{bmatrix}$

Table XI

Elements of JGLS
Analysis #4

Variance Estimates From Ordinary Least Squares	
Airframes:	$S_1^2 = .1402$
Avionics:	$S_2^2 = .2396$
Engines:	$S_3^2 = .0649$

Matrix of Simple Correlation Coefficients
For Ordinary Least Squares Residual Vectors

	Airframes	Avionics	Engines
Airframes	1.0000	.3844	.6139
Avionics	.3844	1.0000	.5036
Engines	.6139	.5036	1.0000

Contemporaneous Covariance

Airframes-Avionics:	$S_{12} = .0704$
Airframes-Engines:	$S_{13} = .0585$
Avionics-Engines:	$S_{23} = .0628$

Elements of

$\hat{V}$ $\qquad\qquad$ $\hat{V}^{-1}$

$$D = \begin{bmatrix} .1402 & .0704 & .0585 \\ .0704 & .2396 & .0628 \\ .0585 & .0628 & .0649 \end{bmatrix} \qquad D^{-1} = \begin{bmatrix} 11.5876 & -.8936 & -9.5943 \\ -.8936 & 5.6604 & -4.6720 \\ -9.5943 & -4.6720 & 28.6014 \end{bmatrix}$$

Table XII

JGLS Coefficients
Analysis #1

Type of Cost	OLS Estimated Coefficient	JGLS Estimated Coefficient	t-ratio of JGLS Estimate
Airframes	-.469013	-.471157	-12.80
	.783941	.807380	6.14
	.416076	.388061	2.42
	.580173	.591760	3.34
	-.608023	-.635549	-2.13
Avionics	-20.352315	-19.954529	-9.97
	.887182	.899911	6.10
	3.320458	3.255030	8.71
	-1.197235	-1.200192	-5.71
	-1.925900	-2.023253	-4.67
Engines	3.413712	3.499367	10.71
	.652594	.604657	3.41
	-.651833	-.698822	-4.70
	-.162780	-.158070	-7.34
	-.937022	-1.024315	-3.53

Table XIII

JGLS Coefficients
Analysis #2

Type of Cost	OLS Estimated Coefficient	JGLS Estimated Coefficient	t-ratio of JGLS Estimate
Airframes	.455526	.447298	12.64
	-.037473	-.030986	-4.87
	-14.221773	-16.969212	-2.65
	.006640	.006854	7.86
Avionics	.121589	.118723	4.95
	-24.423455	-23.779787	-4.41
	.056956	.050389	3.33
	-.239384	-.205861	-1.85
	.001322	.001426	3.68
Engines	-128.784430	-109.300010	-3.52
	.049527	.150362	15.86
	-6.170923	-6.247891	-1.42
	40.034522	28.982606	2.32
	4.261417	4.423444	5.13
	12.096858	13.966332	4.58

Table XIV

JGLS Coefficients
Analysis #3

Type of Cost	OLS Estimated Coefficient	JGLS Estimated Coefficient	t-ratio of JGLS Estimate
Airframes	-.430131	-.428983	-15.06
	-1.006188	-1.006398	-3.61
	.394893	.402198	3.17
	1.044991	1.044212	37.63
Engines	-.201026	-.204332	-9.74
	-.772518	-.813508	-3.39
	.585993	.553670	3.59
	-.685595	-.704641	-5.42
	3.517864	3.571852	12.48

Table XV

JGLS Coefficients
Analysis #4

Type of Cost	OLS Estimated Coefficient	JGLS Estimated Coefficient	t-ratio of JGLS Estimate
Airframes	-.230464	-.239408	-3.40
	1.057095	1.101830	8.07
	-.525572	-.618021	-2.34
Avionics	-.271108	-.272520	-3.53
	3.063690	3.218417	6.50
	-2.050794	-2.209939	-4.49
Engines	14.145672	12.978665	5.56
	-.162844	-.172945	-3.87
	-1.082210	-.871035	-2.67
	1.423641	1.126882	3.09

VII. Conclusions of Analysis

The major emphasis of this study is to determine the effect of joint estimation in developing statistical cost estimating relationships. Initially, however, CER's were developed using Ordinary Least Squares for individual equations. Unadjusted cost observations were used whenever possible and adjustments generally made assessed through the explanatory variables included in the CER's.

By developing joint estimates for airframes, avionics, and engine costs, additional information is utilized in formulating CER's with the logical result that statistical uncertainty is reduced. The four analyses accomplished indicate three general areas of improvement: variance of the estimated coefficients, variance of cost (response) observations, and reduction in prediction intervals at a given confidence level.

Throughout the following discussion, emphasis is on joint estimation; CER's, while derived in a standard manner, were developed as the vehicle to demonstrate Joint Generalized Least Squares rather than an end in themselves. As such, valid questions arise concerning assumptions, elimination of observations, and the ability of the equations to accurately predict future costs; however, regardless of the validity of the CER's themselves, the effect of joint estimation techniques can be demonstrated.

Adjustment of Observations

Two concepts concerning the development of CER's for aircraft systems and their respective sub-elements have gained widespread acceptance: adjustment of given cost data to "constant dollars," and the effect of learning within the manufacturing process for the total procurement of a given aircraft system. The first implies use of a multiplicative index reflecting price escalation to place all costs in the same frame of reference; learning is loosely related to the productivity of the manufacturing process, i.e., the more aircraft produced, the less the average unit cost. Although not the major emphasis of this research, the significance of price escalation and the effect of learning are assessed in the development of the Ordinary Least Squares CER's by including a price index and cumulative quantity as explanatory variables.

A price index (Table III, page 36) was included as an explanatory variable in the first three analyses. The multiplicative price index is included in Analysis #2 even though a linear (additive) functional form is used. The estimated coefficients (Table V, page 46) are negative, although one, for the engine CER, is not statistically significant at the 95 percent confidence level. The negative coefficients indicate a tendency for costs to increase over time; however, little can be determined about the nature of this increase from Analysis #2 because of

the nature of the model and the inconsistency of the coefficients, -14.22, -24.42, -6.17, for airframes, avionics, and engines, respectively.

Price escalation as an explanatory variable is more relevant in the CER development for Analyses #1 and #3 (least squares applied to log-linear variables). The functional form of the CER's is then multiplicative with the estimated coefficients as exponents of the explanatory variables. Since costs are adjusted to constant dollars by multiplying each observation by its applicable escalation index, the "expected" estimated coefficient/exponent of the price index variable is minus one. The five price index coefficients estimated in Analyses #1 and #3 are shown in Table XVI. Also shown is the t-ratio for the hypothesis that the coefficient equals minus one.

Table XVI

Price Index Coefficients

Analysis	Type of Cost	Estimated Coefficient	t-ratio ($b_i = -1$)
1	Airframe	-.6080	1.26
	Avionics	-1.9259	-2.06
	Engines	-.9370	.21
3	Airframes	-1.0062	.02
	Engines	-.7725	.94

In only one case, the avionics (subsystem) CER in Analysis #1, can the hypothesis that the coefficient equals minus

one be rejected at the 95 percent confidence level. In this case, subsystem costs had already been manipulated and adjusted using the price index; therefore, the coefficient may reflect the effects of this adjustment as well as the effect of escalation. The result that the other coefficients are not significantly different from minus one implies that cost adjustment prior to application of least squares does, in fact, indicate the effect of price escalation. By including the index as an explanatory variable, however, a more accurate relationship between costs and all explanatory variables can be obtained, since price escalation adjustment to costs may mask effects other than inflated costs over time, such as productivity.

No attempt was made to ascertain learning curves, or to specifically address the effect of learning in the development of CER's. At the same time, cumulative quantity is considered as an explanatory variable in the first three analyses and can be used to indicate the presence of learning. Tables IV through VI (pages 45-47) show that cumulative quantity was generally a significant explanatory variable for airframes and engines, but not for avionics. This fact may be partially explained by noting that while minor airframe and engine modifications occur between models for a given aircraft system, major changes in avionics systems may be incorporated in varying models. That cumulative quantity was significant in several CER's

indicates average cost per unit does decrease during the procurement process; the specific characteristics of this decrease in cost, as in the development of learning curves, is not pursued, and the effects of learning are assumed to be incorporated in cumulative quantity as an explanatory variable.

Variance of Estimated Coefficients

In addition to explaining cost, the specific effect of explanatory variables is of interest for sensitivity analysis. Statistically, the effect of variables is interpreted through their estimated coefficients. Since intervals which contain the true value of the coefficient, at a given confidence level, are constructed using the estimated coefficients, their variance, and the assumption of their normal distribution, a decrease in the variance of the estimates can significantly decrease the confidence interval for a given coefficient.

The effect of Joint Generalized Least Squares in this respect is shown in Table XVII. The amount of improvement is dependent on the number of equations assumed to interact and their degree of correlation and contemporaneous covariance (Tables VII-XI, pages 48, 51-54). Analysis #3, with only two equations and a low correlation between residuals (.1377), showed almost no improvement; in Analysis #1, even though the correlation coefficients are relatively

Table XVII

Comparison of Variance
for Estimated Coefficients

Analysis	Variance of OLS Coefficient	Variance of JGLS Coefficient	% Decrease
1	.001501	.001356	9.7
	.019277	.017312	10.2
	.028400	.025629	9.8
	.035500	.031402	11.5
	.096281	.088648	7.9
	4.348653	4.004909	7.9
	.023403	.021776	7.0
	.150838	.139701	7.4
	.048249	.044161	8.5
	.201318	.187522	6.9
	.118799	.106693	10.2
	.034094	.031413	7.9
	.024718	.022145	10.4
	.000493	.000464	5.9
	.090537	.084356	6.8
2	.001303	.001251	4.0
	.000057	.000040	29.8
	44.871691	40.894988	8.9
	8.32×10^{-7}	7.49×10^{-7}	10.0
	.000607	.000574	5.4
	34.664770	29.035378	16.2
	.000309	.000229	25.9
	.017061	.012360	27.6
	1.89×10^{-7}	1.50×10^{-7}	20.6
	1239.153100	962.410636	22.3
	9.21×10^{-5}	8.99×10^{-5}	2.4
	22.517541	19.282525	14.4
	202.943610	156.734903	22.8
	.880054	.743535	15.5
	12.038606	9.304036	22.7

Table XVII (Continued)

Analysis	Variance of OLS Coefficient	Variance of JGLS Coefficient	% Decrease
3	.000815	.000812	.4
	.078280	.077925	.5
	.016421	.016102	1.9
	.000771	.000771	-
	.000445	.000441	.8
	.058318	.057646	1.2
	.024277	.023810	.8
	.017216	.016881	1.9
	.083466	.081891	1.9
4	.005830	.004947	15.1
	.029325	.018619	36.5
	.111926	.069934	37.5
	.006278	.005966	5.0
	.325378	.245038	24.7
	.322301	.241867	25.0
	10.023270	5.356625	46.6
	.002314	.001995	13.8
	.200456	.106217	47.0
	.253604	.133288	47.4

small, the combined effect of the three equations results in a decrease in the variances from 5.9 percent to 11.5 percent. Correlation among equations for the remaining two analyses was substantially higher, as was the percentage improvement in the variance of the estimates.

In the interpretation of the effect of variables through their coefficients, it was noted in Section V that four coefficients had values not significantly different from zero at the 95 percent confidence level. For these cases, the Ordinary Least Squares Coefficients, their t-ratios, the Joint Generalized Least Squares coefficients, and their t-ratios are shown below.

OLS Coefficient	t-ratio	JGLS Coefficient	t-ratio
-.608023	-1.96	-.63549	-2.13
-.239384	-1.83	-.205861	-1.85
-6.170923	-1.30	-6.247891	-1.42
-.525572	-1.57	-.618021	-2.34

The t-ratios increase in every case; in two cases, the increase is sufficient to reject the hypothesis that the coefficient equals zero. Obviously, more valid interpretation is possible for these two coefficients.

Variance of Cost Observations

Under Joint Generalized Least Squares, the variance-covariance matrix of the combined observation matrix is

$$V(Y|X) = \alpha^2 V$$

An estimate of the V matrix is constructed from the results of Ordinary Least Squares (Section III, pages 22-23). The estimate of α^2,

$$\hat{\alpha}^2 = \frac{(Y-XB_J)'V^{-1}(Y-XB_J)}{n-k}$$

is then combined with $\hat{V}$ to calculate the estimated variance-covariance matrix for contemporaneous equations:

$$\hat{V}(Y|X) = \hat{\alpha}^2\hat{V}$$

The elements along the main diagonal, $\hat{\alpha}^2 S_i^2$, represent the Joint Generalized variance estimates for individual equations. The Ordinary Least Squares variance estimates, calculated values for $\hat{\alpha}^2$, and corresponding Joint Generalized variance estimates are shown in Table XVIII.

Table XVIII

Comparison of Variance
for Individual Equations

Analysis/ Equation	OLS Variance (S_i^2)	$\hat{\alpha}^2$	JGLS Variance ($\hat{\alpha}^2 S_i^2$)
1/Airframes	.095618		.089759
1/Avionics	.174401	.938723	.163714
1/Engines	.041392		.038856
2/Airframes	975.70374		955.87744
2/Avionics	418.77937	.979680	410.26977
2/Engines	71.988062		70.52526
3/Airframes	.074951		.074909
3/Engines	.030770	.999439	.030753
4/Airframes	.140201		.136502
4/Avionics	.239599	.973619	.233278
4/Engines	.064853		.063120

Prediction

After least squares application and specification of the parameters of a linear model, two types of estimates are calculated: first, expected value of the dependent variable for specified explanatory variables; and second, prediction of a specific future value of the dependent variable. In a model, $Y = X_oB + E$, where estimates of B have been made using least squares (Ordinary or Joint Generalized), the point estimate for both the expected and actual future value is calculated from $\hat{y} = X_o\hat{B}$, where y represents the future observation, X_o the relevant 1 x k vector of given values for k explanatory variables, and $\hat{B}$ the vector of estimated coefficients (from either Ordinary or Joint Generalized Least Squares). Relying on assumptions of the linear model and least squares, variance of the estimates is used to construct prediction intervals which, in a probabilistic context, will contain the true value being estimated at a given level of confidence (usually 95 percent).

Variance of expected value is calculated from

$$\begin{aligned} \mathrm{Var}(\hat{y}) &= \mathrm{Var}(X_o\hat{B}) \\ &= X_o\left[\hat{\sigma}^2(X'X)^{-1}\right] X_o' \end{aligned}$$

for Ordinary Least Squares (Ref 15:122), and

$$\mathrm{Var}(\hat{y}) = X_o \left[\hat{\alpha}^2(X'V^{-1}X)^{-1}\right] X_o'$$

for Joint Generalized Least Squares (Ref 15:237-238).

In predicting specific future values, the prediction error consists of two parts, one for the sampling error of the least squares estimator and the other for the random error of the future observation. The prediction error, $y-\hat{y}$, then has the following variance:

$$\mathrm{Var}(y-\hat{y}) = \hat{\sigma}^2 + \hat{\sigma}^2 x_o (X'X)^{-1} x_o'$$

for Ordinary Least Squares, and

$$\mathrm{Var}(y-\hat{y}) = \hat{\alpha}^2\ \hat{\sigma}^2 + \hat{\alpha}^2 x_o (X'V^{-1}X)^{-1} x_o'$$

for Joint Generalized Least Squares (Ref 15:123,238). Prediction intervals are calculated from the equations shown below.

Expected Cost: $\hat{y} \pm t_{.025} \left[\hat{\sigma}^2 x_o (X'X)^{-1}\right]^{\frac{1}{2}}$ (OLS)

$\hat{y} \pm t_{.025} \left[\hat{\alpha}^2 x_o (X'V^{-1}X)^{-1} x_o'\right]^{\frac{1}{2}}$ (JGLS)

Predicted Cost: $\hat{y} \pm t_{.025} \left[\hat{\sigma}^2 + \hat{\sigma}^2 x_o (X'X)^{-1} x_o'\right]^{\frac{1}{2}}$ (OLS)

$\hat{y} \pm t_{.025} \left[\hat{\alpha}^2 \hat{\sigma}^2 + \hat{\alpha}^2 x_o (X'V^{-1}X)^{-1} x_o'\right]^{\frac{1}{2}}$ (JGLS)

Obviously, reductions in the variance of the equations and estimated coefficients previously discussed imply a subsequent reduction in prediction intervals at a specified confidence level.

Since interest in the CER's developed is in cost rather than the logarithm of cost, one further calculation

is necessary for log-linear functional forms. The form of the intervals above is a point estimate $\pm$ a scalar. For convenience, denote the scalar z. The antilog of a given interval is, therefore,

$$\left[\frac{\exp(\hat{y})}{\exp(z)},\ \exp(z)\exp(y)\right]$$

However, the assumptions of least squares were made on the log-linear form; the multiplicative error terms for actual cost are thus log-normally distributed with median equal to one and expected value

$$E(E_i) = \exp(\sigma^2/2)$$

where σ^2 is the variance associated with the log-linear model (Ref 1:7-9; Ref 16:44-45). The antilog of a least squares point estimate is therefore an estimate of the median rather than the mean. Adjustment using the equation for expected value above is difficult because σ^2 is estimated and $\exp(\hat{\sigma}^2/2)$ is a biased estimate for $E(E_i)$; although some or all of this bias can be eliminated (Ref 6:464-472), calculations are tedious and changes are small. For purposes of this study, prediction intervals for actual cost are constructed as the antilog of least squares predictions of the log-linear form, and therefore use the median cost rather than the mean.

To demonstrate the effect of joint estimation in prediction, the procurement of the F-4D is used. F-4D

procurement consists of three contracts. Characteristics and specific values of explanatory variables are summarized in Table XIX. Cumulative quantity includes F-4C aircraft; however, Navy aircraft, the F-4B and F-4J, are not included in cumulative quantity because of differences in mission, specific manufacturing requirements, and avionics changes specified by the Air Force. Altogether, 54 point estimates are calculated, 27 for both Ordinary and Joint Generalized Squares. Tables XX through XXIII show the point predictions, associated variance, and 95 percent confidence prediction intervals for expected cost and predicted actual cost.

Effects of Joint Estimation

Application of Joint Generalized Least Squares showed varying degrees of improvement in the four analyses accomplished. The primary factors affecting a decrease in statistical uncertainty are the number of equations with contemporaneous covariance, the simple correlation coefficients among Ordinary Least Squares residuals which ultimately determine the relative weight of covariance terms to the elements of variance, and the functional form, linear or log-linear, of CER's.

Analysis #1. This analysis included only Backfill data. CER's were developed in log-linear form. Simple correlations among residuals were extremely low (Table VIII, page 51); even so, the interaction of three equations

Table XIX

Characteristics/Explanatory Variables
for F-4D Costs

Characteristics

Aircraft		Engines	
AMPR Weight	18432	SFC	1.94
Gross TO Weight	58000	Th. to Wt. Ratio	3.417
AMPR Wt-GTO Wt	59568	Maximum Thrust	17.0
Maximum Speed	1239	Compression Ratio	12.9
Combat Speed	725	Number of Engines	2
Wing Loading	95		

Contract Summary

Contract #	Fiscal Year/ Price Index	Quantity/Cum. Quan.*
1	1964/1.494	52/636
2	1965/1.406	222/858
3	1966/1.311	519/1377

*Cumulative quantity includes procured aircraft of the F-4C series. A total of 584 aircraft were purchased.

(Ref CIR Files, ASD Cost Library)

Table XX

F-4D Prediction Comparison
Analysis #1

Point Estimates (Thousands of dollars)

Contract/ Equation	OLS ln(cost)	OLS cost	JGLS ln(cost)	JGLS cost
1/Airframes	7.391397	1621.970	7.397196	1631.404
1/Avionics	6.462469	640.641	6.476470	649.674
1/Engines	6.257788	522.063	6.280153	533.871
2/Airframes	7.287884	1462.473	7.294712	1472.493
2/Avionics	6.579307	720.098	6.599298	734.579
2/Engines	6.265935	526.334	6.295011	541.862
3/Airframes	7.108550	1222.373	7.116290	1231.871
3/Avionics	6.714120	823.958	6.740842	846.273
3/Engines	6.254484	520.341	6.291890	540.176

Variance

Expected Cost OLS	Expected Cost JGLS	Expected Cost % Change	Predicted Actual Cost OLS	Predicted Actual Cost JGLS	Predicted Actual Cost % Change
.012345	.011459	-7.2	.107963	.101218	-6.2
.028575	.026688	-6.6	.202976	.190402	-6.2
.007627	.007309	-4.2	.049019	.046165	-5.8
.016799	.015560	-7.4	.112417	.105319	-6.3
.036349	.033954	-6.6	.210750	.197668	-6.2
.010273	.009872	-3.9	.051665	.048728	-5.7
.023428	.021649	-7.6	.119046	.111408	-6.4
.047147	.044041	-6.6	.221548	.207755	-6.2
.013719	.013245	-3.5	.055111	.052101	-5.5

Table XX (Continued)

Prediction Intervals for Expected Cost

Prediction Interval OLS	JGLS	Change in Range of Interval/%
(1296.905,2028.511)	(1315.157,2023.679)	-23.066/-3.2
(455.237,901.554)	(466.991,903.820)	-9.488/-2.1
(437.860,622.458)	(449.426,634.182)	+.158/+ .1
(1126.620,1898.466)	(1145.501,1892.776)	-24.551/-3.2
(489.838,1058.597)	(506.182,1066.033)	-8.907/-1.6
(429.150,646.527)	(443.592,661.902)	+1.934/+ .1
(898.238,1663.474)	(916.081,1656.520)	-24.796/-3.2
(531.282,1277.866)	(553.751,1293.320)	-7.015/-1.0
(410.998,659.774)	(428.423,681.080)	+4.882/+2.0

Intervals for Predicted Actual Cost

Prediction Interval OLS	JGLS	Change in Range of Interval/%
(837.123,31[illegible]653)	(859.856,3095.262)	-70.124/-3.0
(257.740,15[illegible]84)	(268.972,1569.223)	-34.393/-2.6
(334.248,815.411)	(346.342,822.940)	-4.565/-1.0
(744.678,2872.149)	(766.194,2829.878)	-63.786/-3.0
(284.746,1821.068)	(299.096,1804.126)	-31.292/-2.0
(333.004,831.904)	(347.385,845.214)	-1.070/- .2
(610.332,2448.169)	(629.165,2411.937)	-55.065/-3.0
(318.256,2133.207)	(336.860,2126.041)	-25.770/-1.4
(324.305,834.878)	(341.104,855.429)	+3.752/+ .7

Table XXI

F-4D Prediction Comparison
Analysis #2

Point Estimates (Millions of dollars)

Contract/ Equation	OLS Estimate	JGLS Estimate
1/Airframes	100.996	104.533
1/Avionics	69.970	69.946
1/Engines	26.605	30.406
2/Airframes	171.368	175.188
2/Avionics	92.789	92.221
2/Engines	52.567	56.517
3/Airframes	288.561	293.566
3/Avionics	131.221	129.741
3/Engines	97.563	101.768

Variance

OLS	JGLS	% Chg	OLS	JGLS	% Chg
75.840749	68.972887	-9.1	1051.54449	1024.85032	-2.5
72.502788	64.645694	-10.8	491.28216	474.91546	-3.3
22.842626	18.908303	-17.2	94.83069	89.43356	-5.7
99.741643	91.553962	-8.2	1075.44538	1047.43140	-2.6
95.306249	85.096224	-10.7	514.08562	495.36599	-3.6
28.003720	23.591863	-15.8	99.99178	94.11712	-5.9
315.143730	300.431745	-4.7	1290.84747	1256.30918	-2.7
224.446870	204.663734	-8.8	643.22624	614.93350	-4.4
50.369863	44.746839	-11.2	122.35792	115.27210	-5.8

Table XXI (Continued)

Prediction Intervals for Expected Cost

Prediction Interval OLS	JGLS	Change in Range of Interval %
(83.404,118.587)	(87.757,121.309)	-1.631/-4.6
(52.761,87.179)	(53.697,86.195)	-1.918/-5.6
(16.936,36.274)	(21.609,39.203)	-1.744/-9.0
(151.194,191.542)	(155.860,194.516)	-1.692/-4.2
(73.059,112.519)	(73.578,110.864)	-2.174/-5.5
(41.862,63.272)	(46.691,66.343)	-1.758/-8.2
(252.701,324.421)	(258.553,328.579)	-1.694/-2.4
(100.943,161.499)	(100.828,158.654)	-2.730/-4.5
(83.205,111.921)	(88.236,115.300)	-1.650/-5.7

Intervals for Predicted Actual Cost

Prediction Interval OLS	JGLS	Change in Range of Interval %
(35.492,166.500)	(39.866,169.200)	-1.673/-1.3
(25.175,114.765)	(25.903,113.989)	-1.504/-1.7
(6.905,46.305)	(11.275,49.537)	-1.137/-2.9
(105.124,237.612)	(109.813,240.563)	-1.737/-1.3
(46.966,138.612)	(47.240,137.202)	-1.684/-1.8
(32.338,72.796)	(36.891,76.143)	-1.206/-3.0
(215.986,361.136)	(221.968,365.164)	-1.954/-1.3
(79.965,182.477)	(79.625,179.857)	-2.280/-2.2
(75.185,119.941)	(80.048,123.488)	-1.316/-2.9

Table XXII

F-4D Prediction Comparison
Analysis #3

Point Estimates (Thousands of dollars)

Contract/ Equation	OLS ln(cost)	OLS cost	JGLS ln(cost)	JGLS cost
1/Airframes	7.083216	1191.795	7.082891	1191.408
1/Engines	6.165792	476.178	6.172376	479.324
2/Airframes	7.015516	1113.781	7.015548	1113.816
2/Engines	6.152502	469.892	6.160585	473.705
3/Airframes	6.882431	974.993	6.883020	975.568
3/Engines	6.111449	450.992	6.120836	455.245

Variance

Expected Cost			Predicted Actual Cost		
OLS	JGLS	% Change	OLS	JGLS	% Change
.009450	.009426	-.3	.084401	.084335	-.1
.005027	.005007	-.4	.035797	.035760	-.1
.012867	.012827	-.3	.087818	.087736	-.1
.006691	.006660	-.5	.037461	.037413	-.1
.017856	.017789	-.4	.092807	.092698	-.1
.008855	.008815	-.5	.039625	.039568	-.1

Table XXII (continued)

Prediction Intervals for Expected Cost

Prediction Interval OLS	JGLS	Change in Range of Interval/%
(980.073,1449.255)	(979.998,1448.424)	-.756/-.1
(412.843,549.229)	(415.689,552.701)	+.626/+.4
(886.507,1399.321)	(886.850,1398.869)	-.795/-.2
(398.554,553.999)	(401.941,558.282)	+.896/+.6
(745.141,1275.746)	(745.957,1275.854)	-.708/-.1
(373.166,545.049)	(376.847,549.953)	+1.223/+.7

Intervals for Predicted Actual Cost

Prediction Interval OLS	JGLS	Change in Range of Interval/%
(664.273,2138.242)	(664.209,2137.059)	-1.119/-.1
(325.359,696.908)	(327.573,701.374)	+2.252/+.6
(613.560,2021.821)	(613.750,2021.322)	-.689/ 0
(318.267,693.753)	(320.929,699.208)	+2.793/+.7
(528.208,1799.692)	(528.710,1800.105)	-.089/ 0
(302.095,673.278)	(305.032,679.431)	+3.216/+.9

Table XXXIII

F-4D Prediction Comparison
Analysis #4

Point Estimates (Thousands of dollars)

Equation	OLS ln(cost)	OLS cost	JGLS ln(cost)	JGLS cost
Airframes	7.662500	2127.069	7.672453	2148.345
Avionics	6.503033	667.162	6.547426	697.446
Engines	6.208932	497.170	6.144222	466.017

Variance

Expected Cost OLS	Expected Cost JGLS	Expected Cost % Change	Predicted Actual Cost OLS	Predicted Actual Cost JGLS	Predicted Actual Cost % Change
.017034	.016376	-3.9	.157235	.152878	-2.8
.047457	.041846	-11.8	.287056	.275124	-4.2
.020213	.014353	-29.0	.085066	.077473	-8.9

Prediction Intervals for Expected Cost

Prediction Interval OLS	Prediction Interval JGLS	Change in Range of Interval/%
(1607.679,2814.257)	(1632.651,2826.928)	-12.301/ -1.0
(418.113,1064.556)	(449.726,1081.617)	-14.552/ -2.3
(365.710,675.886)	(359.763,603.653)	-66.286/-21.4

Intervals for Predicted Actual Cost

Prediction Interval OLS	Prediction Interval JGLS	Change in Range of Interval/%
(908.637,4979.348)	(928.682,4969.825)	-29.568/ - .7
(211.409,2105.425)	(226.405,2148.502)	+28.081/ +1.5
(264.792,933.480)	(255.445,850.170)	-73.963/-11.1

resulted in reductions in all relevant variances. One coefficient's t-ratio indicating insignificance under Ordinary Least Squares was increased above the critical value; reductions in the variance of observations and estimates ranged from 3.5 percent to 7.6 percent. The range of prediction intervals generally showed only a small percentage decrease, and in the case of engine costs actually increased, although variance decreased; this contradiction is due to the lack of independence between point estimates and variance under the transformation from logarithms to actual cost. The decrease in variance is partially, or totally, offset by an increase in the point estimate for this specific example. It should be noted that additive rather than offsetting effects could occur when Joint Generalized point estimates are less than Ordinary Least Squares estimates.

Analysis #2. The second analysis included the only explicitly linear CER's, and is the only analysis which dealt with total contract costs rather than average unit cost (per contract or per model). Correlations among residuals are relatively high and substantial reductions are realized in the variance of estimated coefficients and expected cost. Percentage decreases in the variance of the observations and predicted actual cost are less due to $\hat{\alpha}^2 = .97968$, representing a 2.0 percent reduction in the variance of the observations, and dominating the

other portion of variance in predicted actual cost. The range of prediction intervals is reduced by approximately one-half the reduction of their respective variance.

Analysis #3. Because the reported costs for airframes and engines in Project Backfill were more reliable than avionics costs, only these two aspects of cost were included as contemporaneous equations. Results indicate improvements that can be expected with minimal interaction between only two equations. Correlation between Ordinary Least Squares residuals are low; adjustment to the variance of observations was insignificant ($\hat{\alpha}^2$ = .999439, indicating a .06 percent decrease). Coefficients were almost unchanged through application of Joint Generalized Least Squares, as were subsequent point estimates. Although all variances decreased somewhat, no decrease was greater than 2.0 percent. Prediction intervals increased in every case for engine costs, partially due to the transformation problem discussed under Analysis #1.

Analysis #4. Costs were analyzed by aircraft model rather than by contract, and only in this analysis were costs adjusted to constant-year dollars. As in Analysis #2, correlation among residuals was high as were reductions in the variance of coefficients and expected cost; likewise, there was less decrease in the variance of the observations and predicted actual cost ($\hat{\alpha}^2$ = .973619).

As in the other two log-linear analyses, decreases in the range of prediction intervals varied with the relative size of Joint Generalized and Ordinary Least Squares point estimates; however, in this analysis, one point estimate, for engine costs, decreased, and the decrease in the range of prediction intervals is substantially higher.

In general, regardless of degree, variance is reduced through the application of Joint Generalized Least Squares. Of the four types of variance considered in each analysis, the largest percentage savings are realized for the variance of coefficients. Next largest are decreases in the variance of expected cost. This occurs because these two variances are the most closely related to the manner in which Joint Generalized Least Squares "allows" interaction--note that

$$\mathrm{Var}\ (\hat{B}_J) = \hat{\alpha}^2 (X'V^{-1}X)^{-1}$$

$$\mathrm{Var}\ (\hat{y}) = \hat{\alpha}^2 X_o (X'V^{-1}X)^{-1} X_o'$$

Observed reductions for the other variances considered, variance of observations and predicted actual cost, are dominated (and in the former case, defined) by

$$\hat{\alpha}^2 \hat{V}$$

or, for each individual equation, by

$$\hat{\alpha}^2 \hat{\sigma}^2$$

For all of the analyses conducted, $\hat{\alpha}^2$ was greater than 93 percent. Finally reductions in the range of prediction intervals is approximately half the corresponding reduction in variance for linear forms, but highly dependent on the relaticn of point estimates for log-linear CER's.

Application of joint estimation techniques obviously represents greater improvement when correlation/interaction is high. Most significant savings can be anticipated in interpretation of coefficients and predicting expected cost. Reduction in prediction intervals may be significant, although for log-linear models, results may vary. Regardless of reductions in uncertainty, however, the logical interaction of sub-elements of cost can be included through Joint Generalized Least Squares, increasing the validity of statistical models.

Implications of a Joint Distribution

Primary interest has been in the general effect of joint estimation in parametric aircraft cost estimating relationships, and specifically to demonstrate the applicability of Joint Generalized Least Squares. Using the joint distribution of costs assumed under this technique, extensions reducing statistical uncertainty are possible. First, it should be noted that consideration of joint as opposed to several independent distributions implies a

significant departure from past procedure, but represents logical assumptions concerning interactions among sub-costs for the same system. Furthermore, normal distributions are assumed in both cases; interdependence of separate costs is the only major change.

CER's are used throughout the systems acquisition process, including source selection, production decisions, and annual reviews. If, at any time during the process, some costs become known while others are still to be estimated, the conditional distribution of unknown costs given others (derived from the joint distribution) can be used to reduce the variance of estimates.

To demonstrate, the results of Analysis #4 will be used to indicate reductions in the variance of expected cost utilizing the conditional distribution. The Joint Generalized CER's for airframe, avionics, and engine costs were used to estimate expected cost for explanatory variables pertaining to the F-4D. The aspects of cost, denoted y_1, y_2, y_3, respectively, under the assumptions of Joint Generalized Least Squares, possess a tri-variate normal distribution with joint density function as follows (Ref 7:48-50):

$$f(y_1, y_2, y_3) = \frac{|R|^{\frac{1}{2}}}{(2\Pi)^{3/2}} \exp -\tfrac{1}{2}\{(Y-U)'R(Y-U)\}$$

where

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \qquad U = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \qquad R = V^{-1}$$

U represents the mean vector for the components of the Y vector, and R is the inverse of the variance-covariance matrix, V, of the vector Y.

If some of the elements of Y are shown, partitions of these vectors/matrices can be made and the conditional distribution derived (Ref 7:62-64):

Define Y_1 as the vector of elements that are unknown

Y_2^* as the vector of elements that are known

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \quad U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \qquad Y^* = \begin{bmatrix} Y_1 \\ Y_2^* \end{bmatrix} \qquad V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$$

The conditional distribution of Y_1 given Y_2^* is multivariate normal with expected value

$$U_1 + V_{12}V_{22}^{-1}\,(Y_2^*-U_2)$$

and variance-covariance matrix

$$V_{11} - V_{12}V_{22}^{-1}V_{21}$$

In analysis #4, the vector of point estimates for expected costs is calculated from the Joint

Generalized Least Squares CER's. These estimates, as linear functions of normally distributed random variables, are also normally distributed and the corresponding variance-covariance matrix, V, for the joint distribution can be calculated from the quadratic forms discussed in Section VII. Specific results of Analysis #4, in logarithmic form, are shown below.

$$\hat{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 7.6725 \\ 6.5474 \\ 6.1442 \end{bmatrix}$$

$$\hat{V} = \begin{bmatrix} .0164 & .0081 & .0070 \\ .0081 & .1418 & .0085 \\ .0070 & .0085 & .0144 \end{bmatrix}$$

Now suppose that one of the elements is known while the others must still be estimated. For instance, assume that engine costs are known, $y_3^* = 5.6731$, and expected cost for airframes and avionics must still be estimated. Partition as follows:

$$Y^* = \begin{bmatrix} Y_1 \\ Y_2^* \end{bmatrix} \quad Y_1 = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \qquad Y_2^* = y_3^* = 5.6731$$

$$V_{11} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} .0164 & .0081 \\ .0081 & .0418 \end{bmatrix}$$

$$V_{12} = \begin{bmatrix} \sigma_{13} \\ \sigma_{23} \end{bmatrix} = \begin{bmatrix} .0070 \\ .0085 \end{bmatrix}$$

$$V_{21} = \begin{bmatrix} \sigma_{31} & \sigma_{32} \end{bmatrix} = \begin{bmatrix} .0070 & .0085 \end{bmatrix}$$

$$V_{22} = \sigma_{33} = .0144 \; ; \; V_{22}^{-1} = 69.6718$$

The vector U is unknown; however, Y represents an unbiased estimate of U and therefore, an unbiased estimate of

$$E(Y_1|Y_2^*) = U_1 + V_{12}V_{22}^{-1}\,(Y_2^*-U_2)$$

is given by

$$\hat{E}(Y_1|Y_2^*) = \hat{Y}_1 + V_{12}V_{22}^{-1}\,(Y_2^*-\hat{Y}_2)$$

$$= \begin{bmatrix} 7.6725 \\ 6.5474 \end{bmatrix} + \begin{bmatrix} .0070 \\ .0085 \end{bmatrix} \begin{bmatrix} 69.6718 \end{bmatrix} \begin{bmatrix} -.4711 \end{bmatrix}$$

$$= \begin{bmatrix} 7.4427 \\ 6.2684 \end{bmatrix}$$

The variance of Y_1 given Y_2^* is calculated from

$$V(Y_1|Y_2^*) = V_{11}V_{22}^{-1}V_{21}$$

$$= \begin{bmatrix} .0164 & .0081 \\ .0081 & .0418 \end{bmatrix} - \begin{bmatrix} .0070 \\ .0085 \end{bmatrix} \begin{bmatrix} 69.6718 \end{bmatrix} \begin{bmatrix} .0070 & .0085 \end{bmatrix}$$

$$= \begin{bmatrix} .0130 & .0040 \\ .0040 & .0368 \end{bmatrix}$$

The existence of positive covariance among costs and the fact that y_3^* is smaller than its original estimate reduce the estimates for both airframes and engines. More importantly, their variance has been reduced:

Equation	OLS Variance	JGLS Variance	Conditional Variance	% Decrease over OLS	% Decrease over JGLS
Airframes	.0170	.0164	.0130	23.5	20.7
Avionics	.0475	.0418	.0368	22.5	12.0

Finally, the range of 95 percent confidence prediction intervals for expected cost is reduced 29 percent over those computed using Joint Generalized Least Squares, and 30 percent over Ordinary Least Squares intervals:

Equation/Technique Applied	Prediction Interval
Airframes/OLS	(1607.679,2814.257)
Airframes/JGLS	(1632.651,2826.928)
Airframes/Conditional	(1336.888,2180.479)
Avionics/OLS	(418.113,1064.556)
Avionics/JGLS	(449.726,1081.617)
Avionics/Conditional	(349.639,796.239)

VIII. Summary and Directions for Further Study

The contemporaneous covariance assumed to exist in the application of Joint Generalized Least Squares enables the analyst to include more information in the development of parametric estimates, thereby reducing statistical variation. The degree of reduction is dependent on the number of contemporaneous equations and the amount of interaction among costs.

In this sense, the potential for improvement by applying Joint Generalized Least Squares is clearly indicated by the four analyses accomplished. The largest general decrease, for all aspects of variance, is recognized in Analysis #4, where three equations had the highest contemporaneous covariance; at the other extreme, Analysis #3 showed a minimal decrease in variance, as could be expected since correlation between the two equations was low. Analysis #1 provides some insight into the importance of additional equations. Correlation of avionics and the other aspects of cost is low; however, decreases in variance for avionics are comparable to airframes and engines, and overall reduction is significantly better than that realized in Analysis #3. The observed decreases in variance, then, are generated by the cumulative effect of all equations interacting; low correlation, while cause for

concern, does not necessarily indicate a lack of improvement in joint estimation.

In Tables XX, XXII, and XXIII, prediction intervals are not consistently reduced in size for analyses with log-linear forms, although relevant variances decrease. This is due to the nature of the transformation to actual cost: the "uncertainty term" (the scalar z discussed on page 69) is not independent of the point estimate; since point estimates are calculated from different CER's, the relation of the Ordinary Least Squares estimate to the Joint Generalized Least Squares estimate may vary. In the predictions made, Joint Generalized point estimates were generally larger than Ordinary Least Squares estimates; the reverse may be true for other examples.

On the basis of predictive ability, the actual costs for the F-4D were consistently lower than the calculated point estimates; actual costs did, however, fall within the prediction intervals. High point estimates may have been caused by mis-specification of explanatory variables, i.e., not including Navy F-4's in cumulative quantity, and ambiguity in the available listing of aircraft performance characteristics.

The analyses accomplished on the data indicate the applicability of joint estimation techniques to aircraft cost estimation. Functional forms, except for Analysis #2, are consistent with current CER's being used; decreases

in statistical uncertainty were attained in every case. As defense acquisition philosophy tends toward greater trade-offs within a weapons system and includes cost, it seems reasonable to assume that interaction between the aspects of cost will increase. The use of joint estimation techniques can then be of significant value in decreasing statistical uncertainty.

Directions for Further Study

The joint estimation and distribution of variables has been applied here to three sub-elements of aircraft cost. The hypothesized contemporaneous covariance was among elements of total cost for the same contract or model aircraft. Other possible interactions exist within each of the aspects of cost investigated, as well as among other aggregations of the elements of aircraft cost. For example, RAND Report, R-761-PR (Ref 12) ultimately estimates airframe costs from a combination of estimates of engineering labor hours, manufacturing labor hours, tooling hours, and material costs, all of which are related, but are assumed independent. In another report/study (Ref 8), avionics development costs are estimated using an explanatory variable which is itself estimated using least squares techniques; this variable is state-of-the-art (SOA). It can be hypothesized, however, that SOA and avionics

costs are jointly distributed random variables, and not explicitly functionally related.

Specific directions for further study generally will be in areas where logical interaction among variables exists. Extensions of this report include refinement of data, specific investigation of learning, interaction of other aggregations of systems cost, more extensive treatment of conditional distributions, and alternative forms of interaction between functional relationships.

Bibliography

1. Aitchison, J. and J. A. C. Brown. *The Lognormal Distribution.* London: Cambridge University Press, 1966.

2. Alexander, Arthur J. and J. R. Nelson. *Measuring Technological Change: Aircraft Turbine Engines.* RAND Report R-1017-ARPA/PR. Santa Monica: The RAND Corporation, June 1972.

3. Asher, Geraldine, *et. al. Project Backfill.* Unpublished historical cost summary for 12 U.S. Air Force aircraft, sponsored by OSD/SA. Wright-Patterson AFB, Ohio: ASD/ACCI Cost Library, 1965-1966.

4. Campbell, H. G. *Aerospace Price Indexes.* RAND Report R-568-PR. Santa Monica: The RAND Corporation, December 1970.

5. Dreyfuss, D. J., H. E. Boren, Jr., and G. W. Corwin. *Cost Estimating Relationships for Aircraft Airframes.* RAND Working Note WN-8729-PR. Santa Monica: The RAND Corporation, June 1974.

6. Goldberger, A. S. "The Interpretation and Estimation of Cobb-Douglas Functions." *Econometrica.* 35: 464-472 (July-October 1968).

7. Graybill, Franklin A. *An Introduction to Linear Statistical Models.* Volume I. New York: McGraw-Hill Book Company, Inc., 1961.

8. Grimm, Richard W. "Predicting Total Life Costs of Developmental Avionics." *The Proceedings of the IEEE 1974 National Aerospace and Electronics Conference (NAECON '74).* 506-513. New York: Institute of Electrical and Electronic Engineers, Inc., 1974.

9. *Jane's All the World's Aircraft, 1965-1966.* Edited by John W. R. Taylor. New York: McGraw-Hill Book Company, 1966.

10. Johnston, J. *Econometric Methods.* (Second Edition). New York: McGraw-Hill Book Company, Inc., 1972.

11. Kmenta, Jan and Roy F. Gilbert. "Small Sample Properties of Alternative Estimators of Seemingly Unrelated Regressions." *Journal of the American Statistical Association.* 63: 1180-1200 (December 1968).

12. Levenson, G. S., H. E. Boren, Jr., D. P. Tihansky, and F. Timson. *Cost Estimating Relationships for Aircraft Airframes*. RAND Report R-761-PR. Santa Monica: The RAND Corporation, December 1971.

13. Mendenhall, William and Richard L. Schaeffer. *Mathematical Statistics with Applications*. North Scituate, Massachusetts: Duxbury Press, 1973.

14. Summers, Robert. *Cost Estimates as Predictors of Actual Weapon Costs: A Study of Major Hardware Articles*. RAND Memorandum RM-3061-PR. Santa Monica: The RAND Corporation, March 1965.

15. Theil, Henri. *Principles of Econometrics*. New York: John Wiley & Sons, Inc., 1971.

16. Timson, F. S. and D. F. Tihansky. *Confidence in Estimated Airframe Costs: Uncertainty Assessment in Aggregate Predictions*. RAND Report R-903PR. Santa Monica: The RAND Corporation, October 1972.

17. Weinburg, Garrett and Mary Lee Rech. *Aircraft Installed Airborne Equipment Configuration and Cost Data*. Volumes I and II. RMC Final Report UR-097. Bethesda: The Resource Management Corporation, April 1970. (C)

18. Zellner, Arnold. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias." *Journal of the American Statistical Association*. 57: 348-368 (June 1962).

19. Zellner, Arnold. "Estimators for Seemingly Unrelated Regression Equations: Some Exact Finite Sample Results." *Journal of the American Statistical Association*. 58: 977-992 (December 1963).

20. Zellner, Arnold and D. S. Huang. "Further Properties of Efficient Estimators for Seemingly Unrelated Regression Equations." *International Economic Review*. 3: 300-313 (September 1962).

VITA

Patrick William O'Brien was born [redacted] He graduated from high school in Poland, Ohio, in 1962, and attended the United States Air Force Academy from which he received the degree of Bachelor of Science and a commission in the USAF in 1966. He received his pilot's wings in September, 1967. He served as a fighter pilot in Southeast Asia in Cam Ranh Bay AB, RVN, Ubon RTAFB, Thailand, and Udorn RTAFB, Thailand. He entered the Air Force Institute of Technology in June, 1973, arriving from the 31st Tactical Fighter Wing, Homestead AFB, Florida.